Multiscale modeling and simulation of deformation and failure mechanisms of hierarchical alpha-helical protein materials

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEC 10 2009
LIBRARIES

by

Jeremie Bertaud

Ingénieur de l'Ecole Polytechnique (2007)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Civil and Environmental Engineering

at the

ARCHIVES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

© 2009 Massachusetts Institute of Technology
All rights reserved.

Signature of Author.....
Department of Civil and Environmental Engineering
August 7, 2009

Certified by.....
Markus J. Buehler
Associate Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by.....
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Multiscale modeling and simulation of deformation and failure mechanisms of hierarchical alpha-helical protein materials

by

Jeremie Bertaud

Submitted to the Department of Civil and Environmental Engineering
on August 7, 2009 in partial fulfillment of the requirements for the degree of
Master of Science in Civil and Environmental Engineering

Abstract: Alpha-helical (AH) protein structures are critical building blocks of life, representing the key constituents of biological materials such as cells, hair, hoof and wool, where they assemble to form hierarchical structures. AHs play an important mechanical role in biological processes such as mechanotransduction, cell mechanics, tissue mechanics and remodeling. Whereas the mechanics of engineered materials has been widely investigated, the deformation and failure mechanisms of biological protein materials remain largely unknown, partly due to a lack of understanding of how individual protein building blocks respond to mechanical load and how the hierarchical features participate in the function of the overall biological system. In this Thesis, we develop, calibrate, validate and apply two computational models to predict the elasticity, deformation, strength and failure mechanisms of AH protein arrangements and eukaryotic cells over multiple orders of magnitude in time- and length-scales. Our AH protein model is based on the formulation of tensile double-well mesoscale potentials and intermolecular adhesion Lennard-Jones potentials derived directly from results of full atomistic simulations. We report a systematic analysis of the influence of key parameters on the strength properties and deformation mechanisms, including structural and chemical parameters, and compare it with theoretical strength models. We find a weakening effect as the length of AH proteins increases, followed by an asymptotic regime in which the strength remains constant. We also show that interprotein sliding is a dominating mechanism that persists for a variety of geometries and realistic biologically occurring amino acid sequences. The model reported here is generally applicable to other protein filaments that feature a serial array of domains that unfold under applied strain. Although simple, our coarse-grained cell model agrees well with experiments and illustrates how the multiscale approach developed here can be used to describe more complex biological structures. We further show that cytoskeletal intermediate filaments contribute to cell stiffness and deformation and thus play a significant role to maintain cell structural integrity in response to stress. These studies lay the foundation to improve our understanding of pathological pathways linked to AH proteins such as muscular dystrophies.

Thesis Supervisor: Markus J. Buehler
Title: Associate Professor of Civil and Environmental Engineering

Acknowledgements

I would like to express my deep gratitude to Professor Markus J. Buehler for giving me the opportunity to work in his research group during my Master of Science at MIT. I am grateful to him for his continuous support and dynamism. Working with him was a very rewarding experience.

Moreover, I would like to sincerely thank the two MIT undergraduates Joshua Hester and Daniel Jimenez for their contributions to our research projects. It was a great pleasure to supervise them and work together.

For their friendship, their support and collaboration, I would like to express my gratitude to my lab fellows Sinan Keten, Dipanjan Sen, Theodor Ackbarow, Zhao Qin, Sebastien Uzel, Steve Cranford, Andre Garcia, Zhiping Xu, Raffaella Paparcone and Britni Ihle. The excellent atmosphere of the group made my stay a very pleasant experience.

Primary funding for my research project was provided by the Air Force Office of Scientific Research (AFOSR). Additional support was provided from the MIT Schoettler Scholarship. This support is gratefully acknowledged. Additional support from NSF, the Office of Naval Research (ONR) and DARPA is greatly appreciated.

Parts of this research has been submitted for publication and presented at conferences:

Peer reviewed journal publications:

1. Bertaud, J., Buehler, M.J., *Intermediate filament deficient cell show mechanical softening at large deformation,* in submission.

2. Bertaud, J., Hester, J., Jimenez, D.D., and Buehler, M.J., *Energy landscape, structure and rate effects on strength properties of alpha-helical proteins*, in submission.

3. Bertaud, J., Qin, Z. and Buehler, M.J., *Atomistically informed mesoscale model of alpha-helical protein domains.* International Journal for Multiscale Computational Engineering. **7**(3), pp. 237-250, 2009.

4. Bertaud, J., Qin, Z. and Buehler, M.J., *Amino acid sequence dependence of nanoscale deformation mechanisms in alpha-helical protein filaments.* Journal of Strain Analysis for Engineering Design. **44**, doi: 10.1243/03093247JSA533, 2009.

Book chapter:

1. Keten, S., Bertaud, J., Sen, D., Xu, Z., Ackbarow, T. and Buehler, M.J., *Trends in Computational Nanomechanics: Transcending Length and Time Scales.* Springer (New York), in the series *Challenges and Advances in Computational Chemistry and Physics,* in press.

Conference presentations:

Oral presentation

1. Bertaud, J., Hester, J. and Buehler, M.J., *Mesoscale Modeling of Alpha-helical Protein Domains: Size Dependence of Strength and Elasticity*, 2008 Materials Research Society Fall Meeting, December 1 - 5, Boston, MA.

Poster presentation

1. Bertaud, J., Qin, Z. and Buehler, M.J., *Mechanical characterization of alpha-helical protein arrangements using mesoscale modeling techniques*, 2009 National Science Fundation Civil, Mechanical and Manufacturing Innovation Conference, June 22-25, Honolulu, HI.

Table of contents

1 Introduction

1.1 Context

Protein molecules provide the structural basis for critical building blocks of life, forming biological protein materials (BPMs) such as hair, bone, skin, spider silk or cells, which play an important role in providing essential mechanical functions to biological systems [1-8]. Alpha-helical (AH) protein structures are the key constituents of biological materials such as cells, hair, hoof and wool (Figure 1.1), where they assemble to form hierarchical filamentous structures, also called hierarchical alpha-helical protein materials (see left region of Figure 1.1). In other BPMs, one may find key constituents such as beta-sheets (e.g. in spider silk) or collagen proteins (e.g. in bones, tendons, ligaments).

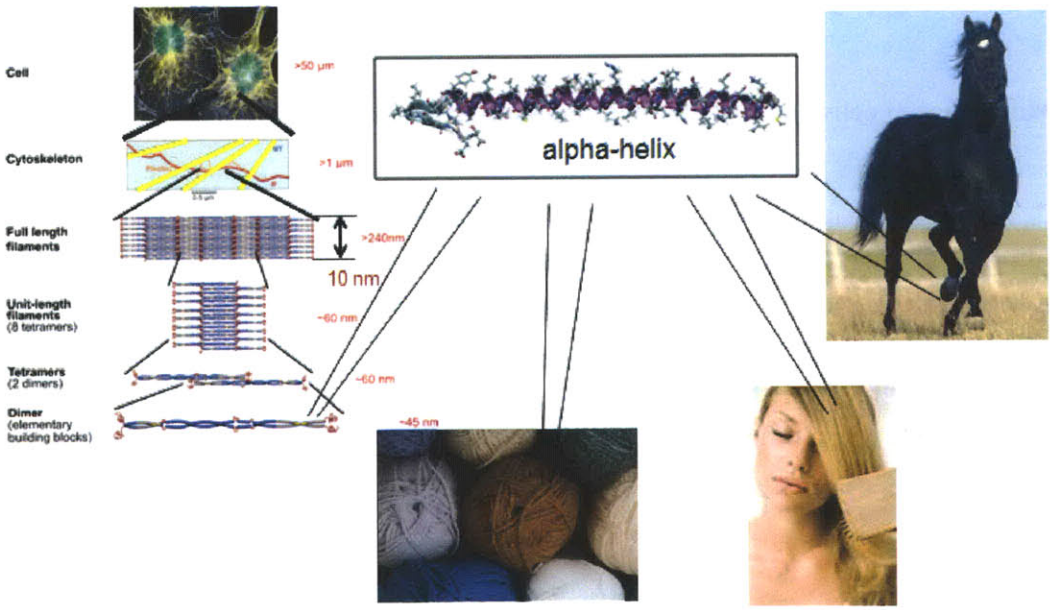

Figure 1.1. The alpha-helical protein structure (top center region) is a key constituent in a variety of biological materials such as cells, hoof, hair and wool. These structures assembly to form hierarchical filamentous structures (left region). The figure in the left part is adapted from [9]. The cell figure (top left region) is reprinted from http://www.ucsf.edu/science-cafe/conversations/wittmann. The wool picture is reprinted from http://www.touk.fr/photos/laine.jpg, the hair picture from http://extensions4hair.com/blog/wp-content/uploads/2009/05/hair_brushing.jpg and the horse picture from http://homeusers.brutele.be/carolo-test/images/quarter_horse.jpg.

Although, as discussed below, the mechanical properties of biological materials have wide ranging implications not only for biology but also for many other fields, the fundamental deformation and failure mechanisms of BPMs remain largely unknown, partly due to a lack of understanding of how individual protein building blocks respond to mechanical load and how these blocks and the hierarchical features participate in the function of the overall biological system.

The mechanical properties of biological materials play an important role, first and foremost in biology. In cells for instance, mechanical sensing is used to transmit signals from the environment to the cell nucleus (e.g. mechanotransduction) or to control tissue formation and regeneration (e.g. tissue modeling or remodeling) [1-5, 10, 11]. Moreover, the structural integrity and shape of cells is controlled by the cell's cytoskeleton, which is composed of hierarchical protein structures. Biological materials such as bone and collagen, which provide structure to our body, or spider silk, used to trap prey, have incredible elasticity, strength and robustness that are unmatched by many synthetic materials. All of these properties are mainly attributed to the structural formation with molecular precision and the combination of nanostructural and hierarchical features universally found in BPMs. Due to these outstanding mechanical properties, biological materials become appealing to other fields in science and engineering. Being able to transfer these biological material concepts into technological applications and new materials design could lead to the development of new materials that mimic or exceed the properties found in biological analogs. However, these applications with potential big payoff in various fields still remain big challenges. In medicine, the characterization of material properties for biological protein materials may also play a crucial role in developing a better understanding of diseases. As pointed out in [8], injuries and genetic diseases are often caused by structural changes in protein materials (e.g. defects, flaws, changes to the molecular structure), resulting in failure of the material's intended function. For instance, point mutations of genes that code certain AH proteins can affect the structure of the protein-based materials such as nucleus envelopes (rapid aging disease *progeria*) and muscle cells (muscle dystrophy disease), causing material weaknesses and thus functional disorders. The approach of characterizing material properties of BPMs enables one to probe how mutations in structure alter the properties of protein materials. In the case of *osteogenesis imperfecta* (brittle bone disease), for instance, molecular-scale models predict a softening of bone's basic collagen constituent [12]. These observations may eventually provide explanations to the molecular origin of certain diseases. Additionally, these findings

provide evidence that material properties play an essential role in biological systems, and that the current paradigm of focusing on biochemistry alone as the cause of diseases is insufficient. It is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties rather than focusing on symptomatic chemical readings alone [8]. Such approaches have already been explored for cancer and malaria, for instance [13, 14].

As pointed out in [8], significant advances in experimental, theoretical and computational materials science have enabled a deeper understanding of BPMs through the linking of structure-process-property (SPP). The material properties of biological materials have been the focal point of extensive studies over the past decades, leading to formation of a research field that connects biology and materials science, referred to as materiomics [6, 8, 15]. Materiomics utilizes mechanistic insight, based on SPP relations in their biological context, to provide a basis for understanding disease processes, to develop new approaches to treating genetic and infectious diseases, injury and trauma as well as to enhance engineered materials via translating material concepts from biology. However, in order to understand deformation and fracture mechanisms of BPMs, it is crucial to elucidate atomistic and molecular mechanisms at each scale. Computational multiscale approaches thereby play a crucial role in transcending through multiple scales in length and time.

As pointed out in [8], the behavior of materials, in particular their mechanical properties, is intimately linked to the atomic microstructure of the material. Whereas crystalline materials show mechanisms such as dislocation spreading or crack extension [16-18], biological materials feature molecular unfolding or sliding, with a particular significance of rupture of chemical bonds such as hydrogen bonds (H-bonds), covalent cross-links or intermolecular entanglement. Additional mechanisms operate at larger length scales, where the interaction of extracellular materials with cells and of cells with one another, different tissue types and the influence of tissue remodeling (at longer timescales) become more evident. The dominance of specific mechanisms is controlled by geometrical parameters, the chemical nature of the molecular interactions, as well as the structural arrangement of the protein elementary building blocks, across many hierarchical scales, from nano to macro (see Figure 1.2).

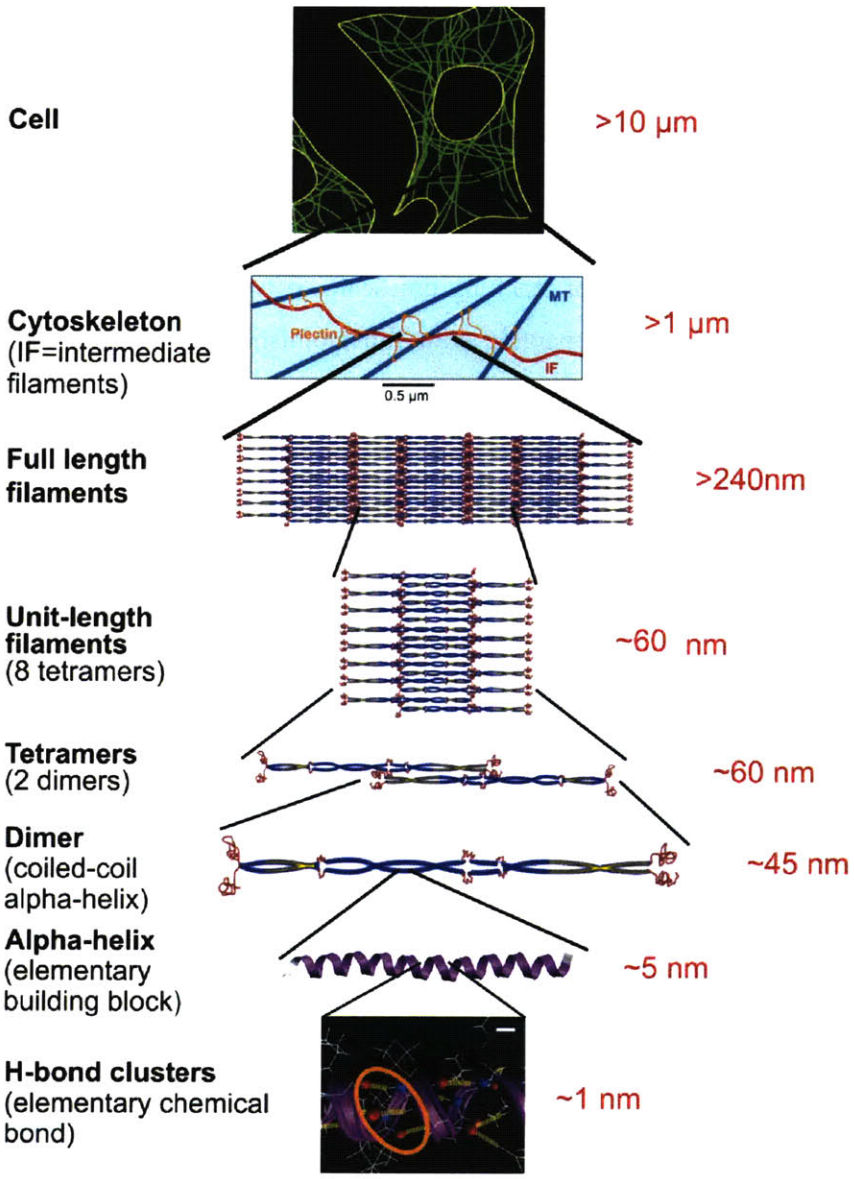

Figure 1.2. Overview over different material scales, from nano to macro, here exemplified for the cellular protein network vimentin intermediate filaments (IFs). In order to understand their deformation and fracture mechanisms, it is crucial to elucidate atomistic and molecular mechanisms at each scale. Computational multiscale approaches thereby play a crucial role in transcending through multiple scales in length and time. Figure reprinted from [9].

The multiscale understanding of how molecular structures participate in macroscale deformation of biological tissues remains an outstanding challenge, and multiscale computational approaches are believed to play a crucial role in advancing this field [6, 8, 15].

Moreover, full atomistic descriptions of proteins are computationally expensive and prohibit simulations at low rates or simulations of very large systems (which includes large protein assemblies). Therefore, multiscale simulation approaches have become increasingly popular in recent years, used to investigate the materiome of biological protein materials through elucidation of SPP relationships [15], and has enabled the direct link between experiment and theoretical bottom-up descriptions of materials (Figure 1.3). Coarse-grained or mesoscale models enable one to simulate the dynamics of large systems over a large range of length- and time-scales. Models are capable of reaching timescales of several microseconds and longer with a quantitative accuracy comparable with full atomistic MD simulations. In the next section, we present the research objectives of the Thesis.

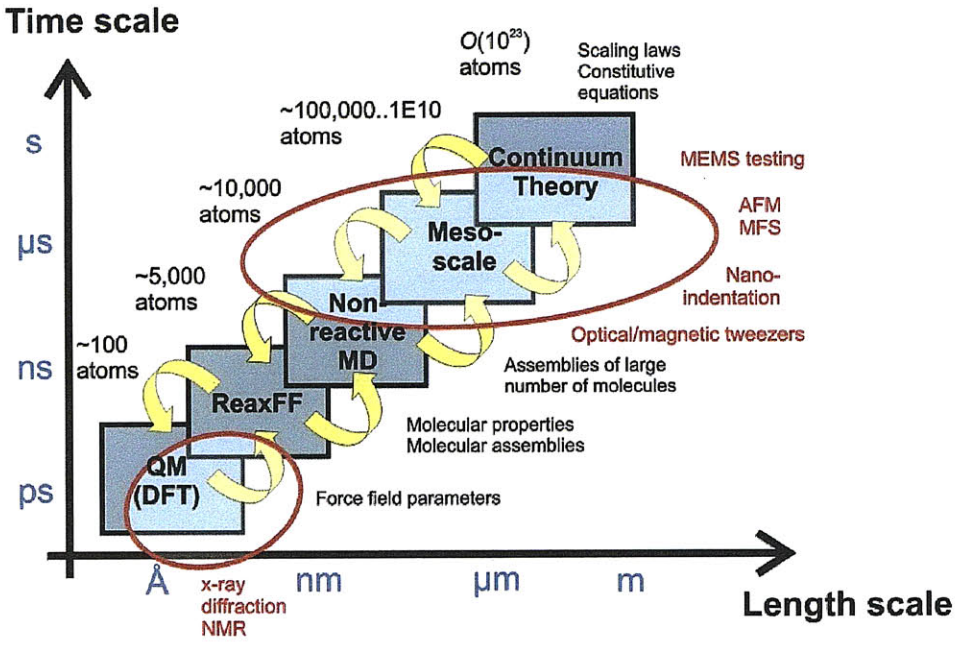

Figure 1.3. Overview over computational and experimental methods. Hierarchical coupling of different computational tools can be used to traverse throughout a wide range of length- and time-scales. Such methods enable one to provide a fundamental insight into deformation and fracture phenomena, across various time- and length-scales. Handshaking between different methods enables one to transport information from one scale to another. Eventually, results of atomistic, molecular or mesoscale simulation may feed into constitutive equations or continuum models. While continuum mechanical theories have been very successful for crystalline materials, biological materials require statistical theories (e.g. the Bell model discussed in Section 3.1 and related formulations) to describe elasticity and strength. Experimental techniques such as the Atomic Force Microscope (AFM), Molecular Force Spectroscopy (MFS), nanoindentation or magnetic/optical tweezers now overlap into atomistic and molecular approaches, enabling direct comparison of experiment and simulation. Techniques such as x-ray diffraction, infrared spectroscopy or NMR provide atomic-scale resolution information about the 3D structure of protein molecules and protein assemblies. Figure reprinted from [19].

1.2 Research objectives and approach

The overall research focus of this Thesis is to study the deformation and failure mechanisms of hierarchical alpha-helical protein materials. The goal of this study can be divided into three main research objectives:

- Discover fundamental structure-process-property (SPP) relationships in alpha-helical (AH) protein materials.

  As discussed above, the strength of protein structures is crucial to identify the mechanical role of protein structures in biological processes such as mechanotransduction or tissue mechanics and tissue remodeling [2, 8, 20-24]. In addition to medical and biological applications, a better understanding of alpha-helices and alpha-helix based protein networks and their resulting mechanical properties could lead to the creation of *de novo* synthetic alpha-helix based materials [25, 26]. Whereas the concept of strength has been widely investigated for engineered materials, the strength of basic biological protein building blocks and how it depends on structural parameters such as the chemical bonding and the protein filament length remains poorly understood. In this Thesis, we study these questions using both a mesoscale computational model for AH protein arrangements and theoretical models. The computational model is built in a way so that one can control the structural parameters such as chemical bonding and length and thus measure their effects on properties such as strength.

- Develop multiscale models that predict elasticity, deformation and fracture of hierarchical assemblies of alpha-helical structures.

  Given their hierarchical features, biological structures are promising materials to enable the development of new multiscale approaches and models that can accurately predict their mechanical behaviors, and show the dominant mechanisms that operate at each hierarchy as well as their contributions to the general macroscopic behavior. In particular, as discussed in the previous section, the mechanical properties of BPMs is intimately linked to the atomic microstructure of the material. The multiscale understanding of how molecular structures participate in macroscale deformation of biological tissues remains an outstanding challenge, and multiscale computational approaches are believed to play a

crucial role in advancing this field. Here, we develop, fit, validate and apply two coarse-grained (or mesoscale) computational models that predict elasticity, deformation and fracture of hierarchical assemblies of AH structures. The first one is for short length- and time-scales and describes the mechanical behavior of AH protein arrangements whereas the second one is dedicated to longer scales and describes the mechanical behavior of a cell with a IF network.

- Develop a model that could be used to improve the current understanding of pathological pathways linked to alpha-helical topologies such as muscular dystrophies.

  As discussed in the previous section, the characterization of material properties for biological protein materials may also play a crucial role in developing a better understanding of diseases. Injuries and genetic diseases are often caused by structural changes in protein materials (e.g. defects, flaws, changes to the molecular structure), resulting in failure of the material's intended function. Here, we develop a simple mesoscale cell model to study the contribution of IF filament network for maintaining cell integrity. As pointed out in [8], it is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties rather than focusing on symptomatic chemical readings alone.

1.3 Outline of the Thesis

The approaches used in this Thesis to study the deformation and failure mechanisms of hierarchical AH protein materials are mainly computational and theoretical, although we make sure to take into account existing experimental approaches and data. Chapters 2 and 3 give an overview of each of these two main approaches, respectively. In Chapter 2 we briefly review basic atomistic and molecular simulation approaches, focusing on molecular dynamics simulation, a selection of force fields, and a brief discussion of multiscale approaches through coarse-graining. In Chapter 3, we review the development of theoretical strength models that emphasize the important role that weak interatomic bonds play in defining the strength properties of biological protein materials. In Chapters 4 and 5 we present the results of the computational work carried out within the scope of this Thesis. Based on their length- and time- scales, we divide the hierarchical AH protein materials into two categories (short and

long scales) and study their deformations and failure mechanisms in these two chapters, respectively. Thus, Chapter 4 presents the studies on the mechanics of AH protein arrangements (short scales), and Chapter 5 the studies on the mechanics of cell and IF networks (long scales). Each of these two chapters begins with a presentation of the computational setup, which includes a detailed description of the mesoscale model formulation, the fitting procedure as well as validation. Then in each of the Chapters the applications of their mesoscale models are presented. For AH protein arrangements, Section 4.3 is dedicated to a presentation and discussion of the study of the sensitivity of the computational results on the parameters of the model such as energy landscape or rate parameters. Section 4.4 presents the discussion of computational results obtained to study the shear strength of parallel AH protein arrangements. Section 4.5 is dedicated to a presentation and discussion of computational results obtained to study the length dependence of strength properties of AH proteins in single and parallel arrangements. We conclude in Section 4.6 with a summary and a discussion. For cells and IF networks, Section 5.2 is dedicated to the applications of the mesoscale cell model. It presents and discusses the computational results obtained to study the contribution of IFs to cell mechanical behavior. Finally, we conclude in Chapter 6 with a summary and discussion of the main findings, the limitations of the current work and an outlook to future research.

2 Computational methods: atomistic based multiscale modeling and simulation

In the following sections we briefly review basic atomistic and molecular simulation approaches, focusing on molecular dynamics simulation, a selection of force fields, and a brief discussion of multiscale approaches through coarse-graining.

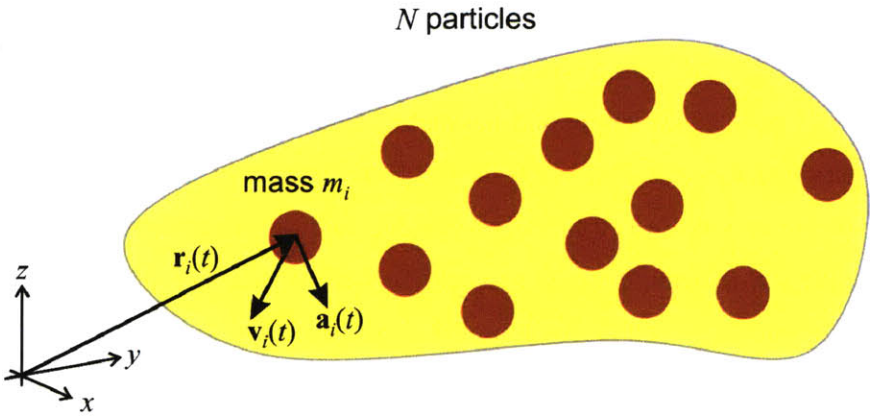

Figure 2.1. Basic formulation of molecular dynamics. This numerical approach predicts the position, velocity and force (acceleration) vector of a set of N particles in a system. Figure reprinted from [19].

2.1 Atomistic simulations

2.1.1 Molecular dynamics formulation

Atomistic molecular dynamics (MD) is a suitable tool for elucidating the atomistic mechanisms that control the deformation and rupture of chemical bonds at nano-scale, and to relate this information to macroscopic materials failure phenomena (see, e.g. review articles and books [19, 27, 28], and recent articles from our group that describes large-scale MD simulation of brittle fracture mechanisms [29-33]). The basic concept behind atomistic simulation via MD is to calculate the dynamical trajectory of each atom in the material, by considering their atomic interaction potentials, by solving each atom's equation of motion

according to $F = ma$, leading to positions $r_i(t)$, velocities $v_i(t)$ and accelerations $a_i(t)$. The basic approach is shown in Figure 2.1.

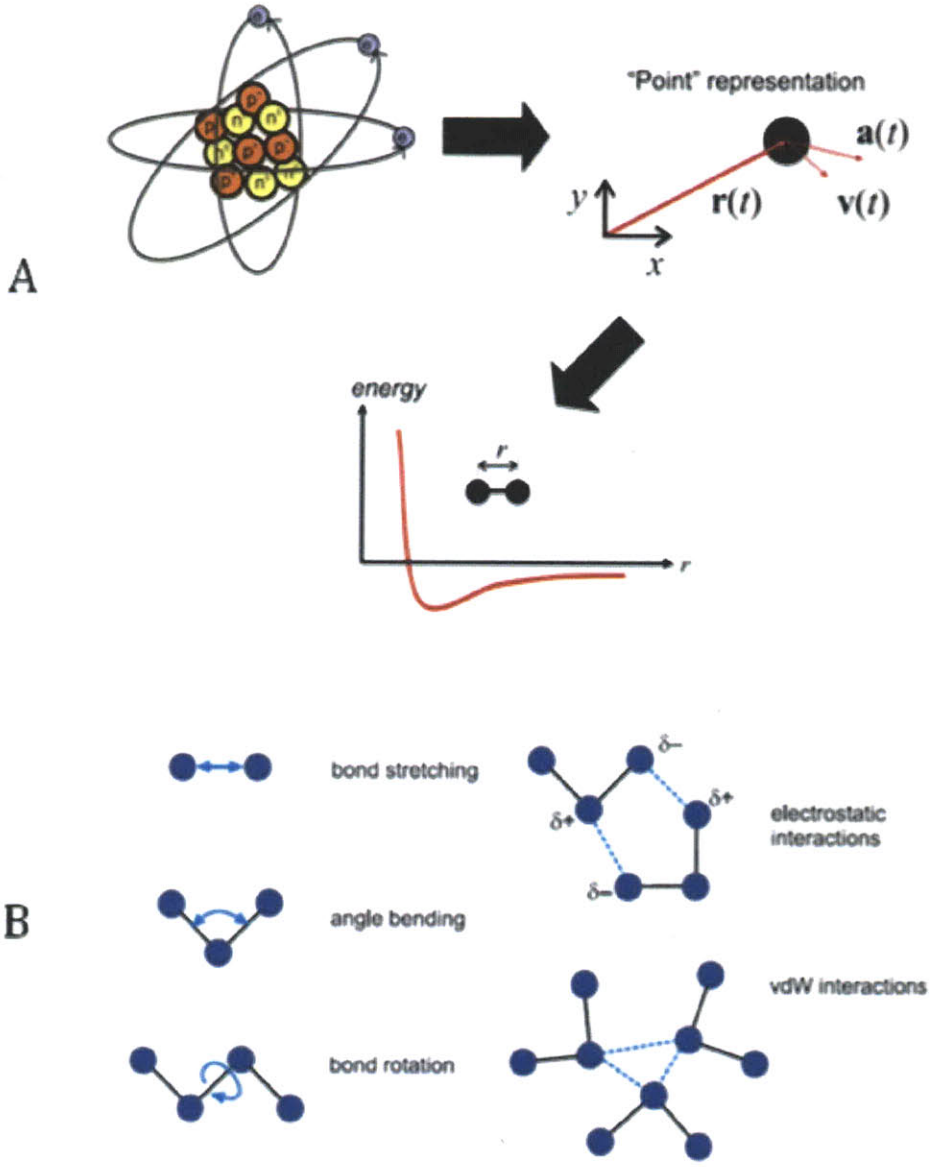

Figure 2.2. Illustration of the concept of point representation as used in molecular dynamics and development of force field models for organic molecules. Panel A: Effective interatomic potentials (as shown in lower part of the figure) provide a description of the energy landscape of the atoms, capturing the overall effect of the interactions of the atoms due to their specific quantum mechanical structure (electrons, neutrons, protons, as shown in the upper left part). Panel B: Illustration of how the chemical bonding characteristics is modeled by decomposing into energy contributions due to bending, stretching, rotation and other interactions (similar as the approach used in the definition of the CHARMM model and other related force fields). Figure adapted from [19].

The numerical integration of Newton's law by considering proper interatomic potentials to obtain interatomic forces enables one to simulate a large ensemble of atoms that represents a larger material volume, albeit typically limited to several nanoseconds of timescale. The availability of interatomic potentials for a specific material (based on the characteristic type of chemical bonding) is often a limiting factor for the applicability of this method.

Classical molecular dynamics generates the trajectories of a large number of particles, interacting with a specific interatomic potential. Thereby, the complex 3D structure of an atom (composed of electrons and a core of neutrons and protons) is approximated by a point particle (Figure 2.2(A)).

Molecular dynamics is an alternative approach to methods like Monte-Carlo, with the distinction that MD provides full dynamical information and deterministic trajectories, which is crucial to describe highly driven phenomena such as catastrophic failure of a protein domain due to laterally applied loads. It is emphasized that Monte-Carlo schemes provide certain advantages as well; however, this point will not be discussed further here as most simulation studies reviewed here are carried out with a MD approach. The total energy of the system is written as the sum of kinetic energy (K) and potential energy (U),

$$E = K + U, \tag{2.1}$$

where the kinetic energy is

$$K = \frac{1}{2} m \sum_{j=1}^{N} v_j^2, \tag{2.2}$$

and the potential energy is a function of the atomic coordinates r_j,

$$U = U(r_j), \tag{2.3}$$

with a properly defined potential energy surface $U(r_j)$. The numerical problem to be solved is a system of coupled second order nonlinear differential equations:

$$m \frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \qquad j = 1..N, \tag{2.4}$$

which can only be solved numerically for more than two particles, $N > 2$. Typically, MD is based on updating schemes that yield new positions from the old positions, velocities and the current accelerations of particles:

$$r_i(t_0 + \Delta t) = -r_i(t_0 - \Delta t) + 2r_i(t_0)\Delta t + a_i(t_0)(\Delta t)^2 + ... \quad (1.5)$$

The forces and accelerations are related by $a_i = f_i / m$. The forces are obtained from the potential energy surface – sometimes also called force field – as

$$F = m\frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \qquad j = 1..N. \quad (1.6)$$

This technique can also be used for not only single atoms but also groups of atoms as in the case of coarse-grained meso-scale approaches. Provided interatomic potentials are available, MD is capable of directly simulating a variety of materials phenomena, for instance the response of an atomic crystal lattice to applied loading under the presence of a crack-like defect, or the deformation mechanisms of biological molecules including nucleic acids and proteins.

One of the strengths and a unique feature of atomistic methods is its very fundamental viewpoint of materials phenomena. The only physical law that is put into the simulations is Newton's law and a definition of how atoms interact with each other. Despite this very simple basis, very complex phenomena can be simulated. Unlike many continuum mechanics approaches, atomistic techniques require no *a priori* assumption on the defect dynamics. Once the atomic interactions are chosen, the complete material behavior is determined. Choosing appropriate models for interatomic interactions provides a rather challenging and crucial step that remains subject of a very active discussion in the scientific community. A variety of different interatomic potentials are used in the studies of biological materials at different scales, and different types of protein structures may require the use of different atomistic models. A drawback of atomistic simulations is the difficulty of analyzing results and the large computational resources necessary to perform the simulations. Due to computational limitations, MD simulations are restricted with respect to the timescales that can be reached, limiting overall time spans in such studies to tens of nanoseconds, or in very long simulation studies to fractions of microseconds. Therefore, many MD simulation results of dynamically

stretching protein molecules, for instance, have been carried out at large deformation rates, exceeding several m/sec.

Recent advances in computational power now enable the simulation of billions of particles in MD simulations, reaching dimensions on the order of micrometers. Figure 2.3 depicts the historical development of computational power over the past decades.

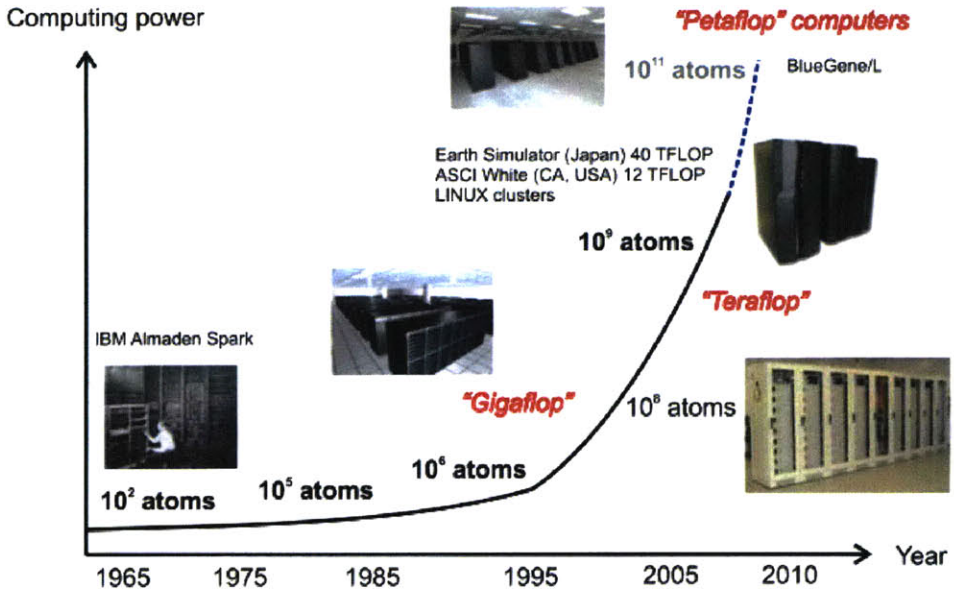

Figure 2.3. Development of computing power over the past decades. The development illustrates the emergence of petaflop computers in the next few years. The plot also summarizes the number of atoms that can be treated with these computing systems; these numbers are developed for simple interatomic potentials with short cutoffs. It is noted that for CHARMM, the number of atoms that can be simulated is significantly smaller. Figure reprinted from [19].

In the next two sections, we provide a brief review of popular interatomic force fields and modeling approaches suitable for simulating the behavior of protein structures. We refer the reader to more extensive review articles for additional information, in particular regarding force field models [34-37].

2.1.2 CHARMM force field

All-atom force fields are predominantly used in molecular dynamics simulations of biological materials at the nanoscale as they generally are the most reliable yet computationally efficient

way of studying dynamics of macromolecules. A wide range of force fields and simulation programs are currently available, most notably the AMBER, the CHARMM force fields and programs, the OPLS force field, the GROMOS/GROMACS [38] packages are commonly used in all-atom molecular dynamics. The NAMD [39] program is a popular code that can carry out computations using CHARMM and other force fields. For the sake of brevity, the main aspects of the CHARMM force field and its implementation in NAMD will be discussed here; the basic concepts of the MD technique and force field formulations are common to all packages used in the field (for a general review, see for instance [35, 40]

The CHARMM force field is widely used in the protein and biophysics community, and provides a reasonable description of the behavior of proteins. The parameters in force fields are often determined from more accurate, quantum chemical simulation models by using the concept of force field training [41] (see the hierarchical coupling schematic shown in Figure 1.3). Parameters for the CHARMM force field have been meticulously optimized and revised over the years taking into consideration a wide variety of input including *ab initio* results, experimental crystal structures and geometries, as well as vibrational spectra [42].

The potential includes bonding and non-bonding (interaction) terms to describe short and long-range forces between particles. In the CHARMM model, the mathematical formulation for the empirical energy function that contains terms for both internal and external interactions has the form:

$$U(\dot{R}) = \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{UB}} K_{UB}(S-S_0)^2 + \sum_{\text{angle}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} \mathrm{K}_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{impropers}} \mathrm{K}_{\text{imp}}(\phi-\phi_0)^2 + \sum_{\text{nonbond}} \varepsilon\left[\left(\frac{R_{\min(i,j)}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min(i,j)}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \tag{2.7}$$

where K_b , K_{UB}, K_θ, K_χ and K_{imp}, are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, δ, and ϕ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium positions for the individual terms. Figure 2.2(B) shows a schematic of the individual energy contributions listed in eq. (2.7).

The Coulomb and Lennard-Jones 6-12 terms constitute the external or nonbonded interactions; ε is the Lennard-Jones well depth and $R_{\min(i,j)}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, ε_1 is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. In the CHARMM force field, no additional terms are used for H-bonds, since the combination of charge and Lennard-Jones contributions were verified to be adequate for describing protein, solvent and interface hydrogen bonding. In all-atom force fields, water molecules are generally also treated explicitly. Parameters of the force field generally are specified considering a specific water model (e.g. TIP3P dimer model for CHARMM) [35, 40].

The CHARMM force field belongs to a class of models with similar descriptions of the interatomic forces; other models include the DREIDING force field [43], the UFF force field (="Universal Force Field") [44], or the AMBER model [34, 45]. In CHARMM and other classical force fields, bonded terms are modeled with harmonic springs or its variations, and therefore cannot be modified (e.g. towards a different chemical state, such as from sp^2 to sp^3) or broken once defined by the connectivity input obtain from the topology of the molecule. Further, the atomic charges are fixed and cannot change during a simulation. These simplifications improve the simulation speed drastically and are not a major issue for most simulations studying conformational changes of proteins under ambient physiological conditions. On the other hand, simulations in extreme conditions such as mechanical perturbations (e.g. protein unfolding studies) or harsh chemical environments require reactive force fields that can take into account changes in fixed charges of the molecules, formation/breaking of new bonds and variations in bond order.

2.2 Review on multiscale methods for protein materials

Multiscale simulation models for protein materials have become increasingly popular in recent years, and have enabled the direct link between experiment and theoretical bottom-up descriptions of materials (Figure 2.4). Here we provide a review of popular multiscale approaches to describe the mechanics of proteins in a multiscale setting.

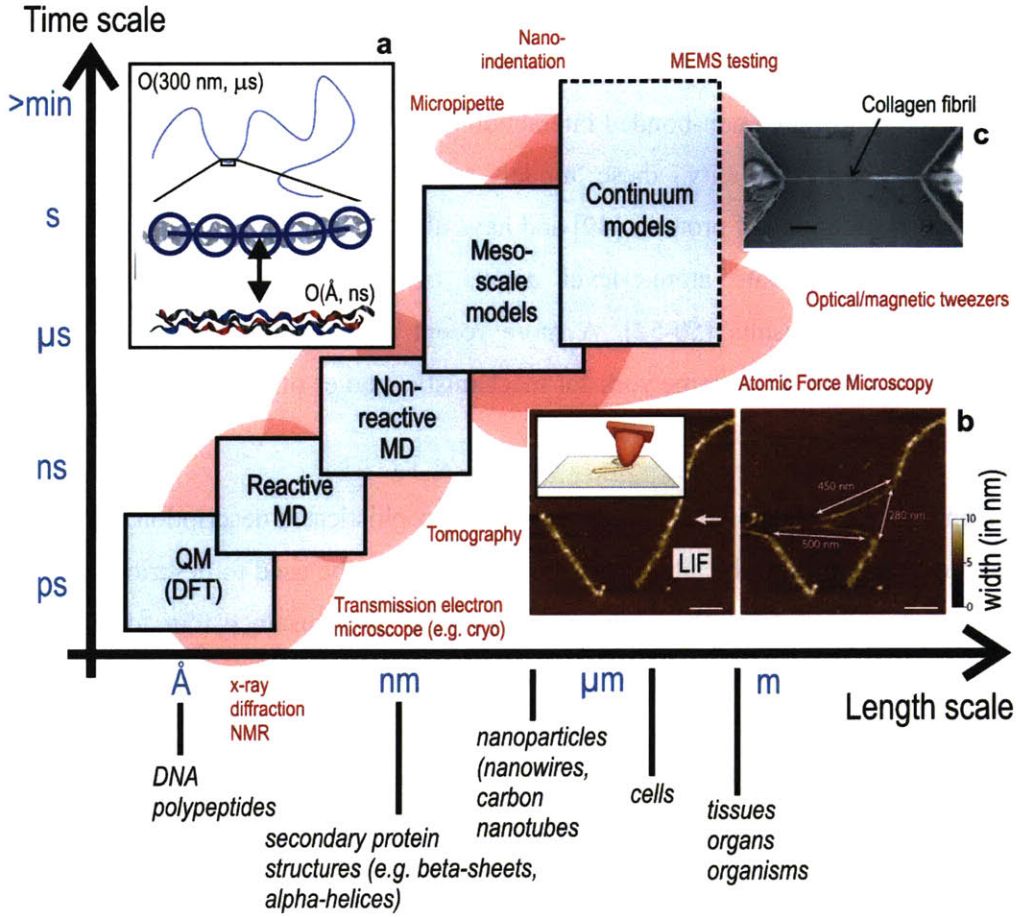

Figure 2.4. Overview over computational and experimental methods. Hierarchical coupling of different computational tools can be used to traverse throughout a wide range of length- and time-scales. Such methods enable one to provide a fundamental insight into deformation and fracture phenomena, across various time- and length-scales. Handshaking between different methods enables one to transport information from one scale to another. Eventually, results of atomistic, molecular or mesoscale simulation may feed into constitutive equations or continuum models. While continuum mechanical theories have been very successful for crystalline materials, biological materials require statistical theories (e.g. the Bell model discussed in Section 3.1 and related formulations) to describe elasticity and strength. Experimental techniques such as the Atomic Force Microscope (AFM), Molecular Force Spectroscopy (MFS), nanoindentation or magnetic/optical tweezers now overlap into atomistic and molecular approaches, enabling direct comparison of experiment and simulation. Techniques such as x-ray diffraction, infrared spectroscopy or NMR provide atomic-scale resolution information about the 3D structure of protein molecules and protein assemblies. Figure reprinted from [8].

Single bead models are the most direct approach taken for studying macromolecules. The term "single bead" derives from the idea of using single beads, that is, point masses, for describing each amino acid in a protein structure. The Elastic Network Model (ENM) [46], Gaussian Network Model (GNM) [47] and Go-model [48] are well known examples that are based on such bead model approximations. These models treat each amino acid as a single bead located at the C_α position with mass equal to the mass of the amino acid. The beads are

connected via harmonic bonding potentials, which represent the covalently bonded protein backbone. In Go-like models, an additional Lennard-Jones based term is included in the potential to describe short-range non-bonded interactions between atoms within a finite cutoff separation. Despite their simplicity, these models have been extremely successful in explaining thermal fluctuations of proteins [49] and have also been implemented to model the unfolding problem to elucidate atomic-level details of deformation and rupture that complement experimental results [50-52]. A more recent direction is coupling of ENM models with a finite element-type framework for mechanistic studies of protein structures and assemblies [53].

Using more than one bead per amino acid provides a more sophisticated description of protein molecules. In the simplest case, the addition of another bead can be used to describe specific side-chain interactions in proteins [54]. Higher level models, for instance four to six bead descriptions, capture more details by explicit or united atom description for backbone carbon atoms, side chains, carboxyl and amino groups of amino acids [55, 56]. Even coarser level multiscale modeling methods have been reported more recently, applied to model biomolecular systems at larger time- and length-scales. These models typically employ superatom descriptions that treat clusters of amino acids as "beads". In such models, the elasticity of the polypeptide chain is captured by simple harmonic or anharmonic (nonlinear) bond and angle terms. These methods are computationally quite efficient and capture shape dependent mechanical phenomena in large biomolecular structures [57], and can also be applied to collagen fibrils in connective tissue [58] as well as mineralized composites such as nascent bone [59]. In Chapter 4, we develop such a coarse-level description of alpha-helical protein domains.

3 Theoretical concepts: strength models for protein materials

The strength of biological protein materials is determined by bonds of different strength and how they interplay in the hierarchical structural arrangement. In particular, weak H-bonds play a crucial role in defining protein constituents' mechanical strength, and their behavior is critical in formulating bottom up strength models. In the following sections, we review the development of theoretical strength models that emphasize the important role that weak interatomic bonds play in defining the strength properties of biological protein materials.

3.1 Strength of a single bond

3.1.1 Bell's model: a force dependent dissociation rate

The Bell's model is a simple and quite popular phenomenological model which describes the frequency of failure of reversible bonds [60]. The concept of "reversibility" thereby means that an individual bond can break under no force if one waits a sufficiently long time, and that it can reform spontaneously. Such bonds may be associated with electrostatic, van der Waals (vdW), or H-bond interactions. The frequency of failure, also called dissociation rate or off rate, is defined as the inverse of the bond lifetime and often used as a concept to discuss the dynamical behavior of such bonds.

Bell's model explains the force dependence of the off rate and thus shows the significant role of mechanical force in biological chemistry. For instance, this theory can be applied to describe the forced unbinding of biological adhesive contacts such as adhesion of cells to cells. Bell's model is an extension of the transition state theory for reactions in gases developed by Eyring and others [61]. Inspired also by Zhurkov's work on the kinetic theory of the strength of solids [62], Bell predicted for the first time that the off rate of a reversible bond, which is the inverse of the bond lifetime, increases when subjected to an external force f. Indeed, the rupture of bonds occurs via thermally assisted crossing of an activation barrier E_b which is reduced by $f \cdot x_b$ as the applied force f increases, x_b being the distance between the bound state and the transition state (see Figure 3.1). Thus the Bell off rate expression is

$$k = \omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B \cdot T}\right), \tag{3.1}$$

where ω_0 is a natural vibration frequency and $k_B \cdot T$ the thermal energy. The force $f_0 = E_b / x_b$ represents the force to vanish completely the energy barrier and gives a very rough value of the rupture force.

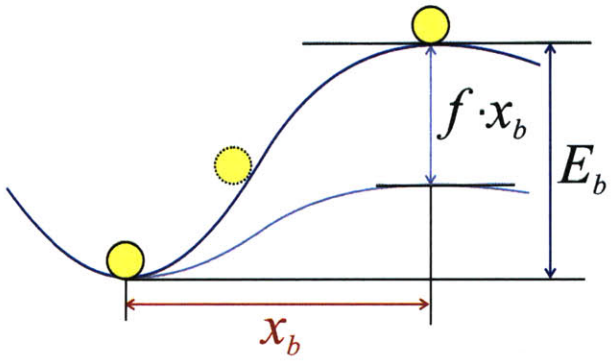

Figure 3.1. Evolution of the energy landscape of a bond subjected to a force. The minimum corresponds to the bound state. The transition state is the peak of the potential and corresponds to the bond rupture. The parameter x_b is the distance between the bound state and the transition state. According Bell's model, the rupture of a bond occurs via thermally assisted crossing of an activation barrier E_b which is reduced by $f \cdot x_b$ as the applied force f increases.

This conjecture was established long before single molecule experiments were performed. Later, it became very successful especially to describe forced unfolding of biological molecules. Indeed, the model enables one to characterize the bonds, their ruptures and their energy landscape profiles from the fitting with experimental or simulation results. Lastly, as discussed in a following section, this theory leads to the development of new theoretical models to explain more complex molecular bonds such as multiple bonds in parallel.

Although successful, the approach has some limitations that have led to several refinements. A few major refinements are reviewed in more details in the next sections. A major limitation is that the Bell's model deals only with constant external force and does not explain loading rate dependence of strength. This limitation is important since we usually cannot apply a constant force in experiments due to the very short thermal impulse time compared with the measurement time [63]. Another limitation is the fact that x_b does not depend on the force. In addition, the multidimensional nature of the energy landscape of biomolecules which can lead to multiple unfolding pathways is not taken into account [64].

3.1.2 Evan's extension: a loading rate dependence of strength

The Evans extension to the Bell model attempts to solve some of the limitations of the Bell's model. It gives a more general relation of the off rate by taking into account force dependent terms such as $x_b(f)$. Moreover it explains for the first time the rate dependence of strength.

The general off rate relation derived by Evans is extended from the Brownian dynamics theory of Kramer. Kramer's theory deals with reactions kinetics in liquids. It gives the thermal noise-driven rate of escape of a particle over a potential barrier using the Fokker-Planck approximation [61, 63]. Evans extended it by including an external pulling force f. Thus he reduced the dimensions of the reaction path to one dimension (coordinate x) and added the potential $-f \cdot x$. Now we briefly present the steps of derivation of the off rate. The time evolution of the probability density of the molecular configuration $p(x,t)$ is described by the Fokker-Planck equation. From the solution, we derive the bond survival probability at time t, referred to $S(t)$, and the mean first passage time, T, which defines the bond lifetime (mean time of passage from the bound state to the unbound state). Thus, we obtain the bond off rate as the inverse of the bond lifetime:

$$k = k_0 g(f) \exp\left(\frac{\Delta E_b(f)}{k_B T}\right) \tag{3.2}$$

where k_0 is a prefactor which contains the Arrhenius dependence on barrier energy scaled by a characteristic time constant, $g(f)$ is a function which depends on deformation of energy landscape by external force, and $\Delta E_b(f)$ is the reduction in energy barrier height. This off rate relation is more general compared with Bell's relation. Indeed, it does not use the linear approximation $\Delta E_b(f) = -f \cdot x_b$ for the reduction in energy barrier height. Thus it may be physically more relevant since it allows the positions of transition state and bound state to change under external force.

The second and probably most important contribution of Evans was to demonstrate that the strength of bonds depends crucially on the loading rate. He theoretically showed that, above a critical loading rate, the force of rupture increases logarithmically with the loading rate, and thus explained what was already observed by many experiments. Now, we briefly present the steps of derivation to obtain the rupture force expression at a given loading rate. The force f applied to the bond increases with time t such as $f = r_f \cdot t$ where r_f is the loading rate.

Because of thermal fluctuations, the rupture events are stochastic and thus give a rupture force density distribution $p(f)$ (see Figure 3.2(a)) . The probability $p(f) = p(t = f/r_f)$ that the bond breaks at time t is the probability that the bond survives to time t, $S(t)$, multiplied by the bond off rate $k(f)$ which is the instantaneous probability that the bond breaks at time $t = f/r_f$ given that it survives to time t:

$$p(f) = p(t = f/r_f) = k(f) \cdot S(t) \tag{3.3}$$

The bond strength f^* is defined as the most probable rupture force and corresponds to the peak of the force density distribution. Thus we can obtain f^* from the following equation

$$\frac{dp(f)}{df}\Big|_{f=f^*} = 0 \tag{3.4}$$

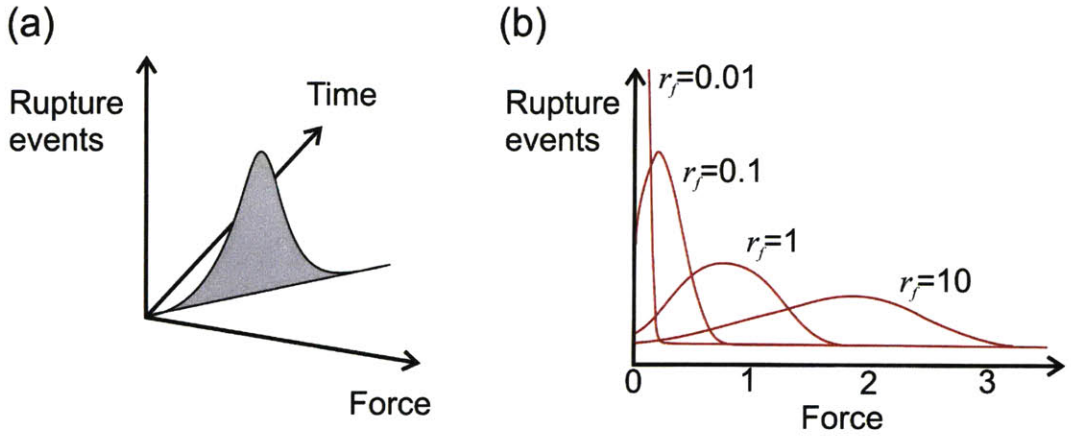

Figure 3.2. Subplot(a) illustrates the rupture forces predicted for bonds over time and force in mechanical probe tests [63]. Loaded by a ramp of force, the off rate increases steadily, but the likelihood of bond survival decreases simultaneously. Thus the frequency of failure can reach a maximum at some time equivalent to force; the peak defines the bond strength f^*. Subplot (b) shows that rupture force distributions shift with loading rate r_f. Figure adapted and redrawn from [63].

Provided that x_b and E_b remain constant, one can demonstrate the logarithmic dependence of strength on the loading rate as follows:

$$f^* = \frac{k_B \cdot T}{x_b} \log\left(\frac{r_f \cdot x_b}{k_B \cdot T \cdot k_0}\right) \text{ for } r_f > r_c \equiv \frac{k_B \cdot T \cdot k_0}{x_b} \tag{3.5}$$

where k_0 is the off rate in the absence of force. For smaller loading, $r_f < r_c$, the most probable rupture force is zero. Figure 3.2(b) illustrates this increase of the most probable rupture force with the loading rate.

In force probe techniques, a transducer is usually moved at constant speed relative to a substrate in which the bond is anchored. The pulling speed v is linked to the loading rate r_f through the stiffness K of the transducer $r_f = K \cdot v$. Thus, this description of rate dependence is very significant for experiments because it enables to get the bond constants k_0 and x_b from a simple linear regression on the force-log(r_f) curve. Moreover it rationalizes the variation among rupture force values obtained from different experimental and simulation techniques which were using different loading rates. However, for a wide range of loading rate we may get a non-linear behavior. An explanation can be that x_b and E_b do not remain constant (change of mechanism, multiple energy barriers), or the elasticity contribution changes with the loading rate.

3.1.3 Other refinements of Bell's model

Several other attempts have been made to extend the very simple Bell's model. For instance, it has been shown that rebinding can have a great impact on strength [65-67]. Similar to a force dependent x_b term, the existence of a rebinding rate can alter the logarithmic rate dependence of rupture force. In non-equilibrium pulling regime, other models attempt also to explain non-logarithmic rate dependence [68]. Moreover, other extensions try to implement the influence of the transducer stiffness in order to explain the disparities in measured unbinding force among different methods [69]. As a last example of extension, we can mention the existence of models which take into account the energy landscape roughness of bonds [64, 70].

3.2 Strength of complex molecular bonds

According Bell's model, the lifetime of an individual weak bond such as an H-bond is very low. However in biology weak bonds can provide more significant strength by forming

arrangement of multiple bonds. In this section, we briefly review theoretical models which study strength of various bond arrangements. A common feature of these models is that they all implement the Bell's model to describe the off rate of a simple bond at a given time *t*, in other words the probability for the bond to break at time *t* given that the bond survives to time *t*.

3.2.1 Multiple bonds in parallel

Many attempts have been made to study the strength of multiple parallel bonds [60, 65, 67, 71, 72]. These models try to describe the influence of various parameters on the strength. For instance, interesting parameters can be the potential profile of a simple bond, the number of bonds, the rebinding rate (neglected, constant, Bell based, Boltzmann based, etc.), the constant external force, the loading rate and the stiffness of the transducers. Figure 3.3 presents an example of geometry with a given set of parameters.

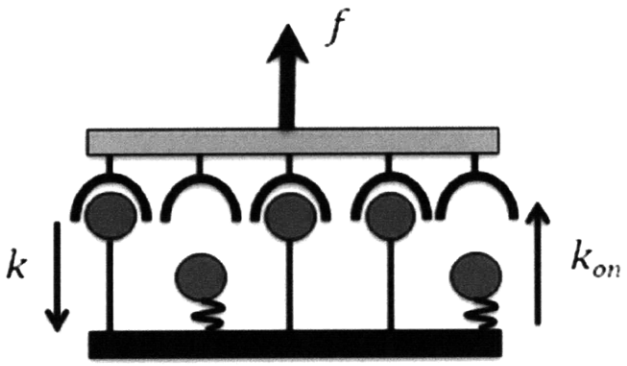

Figure 3.3. Example of a schematic representation of an adhesion cluster under constant force [71]. There are N_0 (= 5) bonds in parallel, of which i (= 3) are closed and equally share the constant dimensionless force f. Single closed bonds rupture with dissociation rate $k = k_0 \exp\left(\frac{f \cdot x_b}{k_B \cdot T \cdot i}\right)$ and single open bonds rebind with force-independent association rate k_{on}. This model has three parameters: cluster size N_0, dimensionless rebinding rate $\gamma = k_{on}/k_0$ and force f. Figure adapted and redrawn from reference [71].

The theoretical description of the rupture of parallel bonds is traditionally achieved by one of the two approaches presented below, the deterministic approach and the stochastic approach. We first discuss the deterministic approach. This method describes the time evolution of the mean number of closed bonds $N(t)$:

$$\frac{dN}{dt} = -N(t) \cdot k_0 \cdot \exp\left[\frac{f \cdot xb}{k_B \cdot T \cdot N(t)}\right] + k_r \cdot \left[N_0 - N(t)\right] \tag{3.6}$$

with N_0 the total number of bonds, f the external force, k_r the rebinding rate of a single broken bond and $k_0 \cdot \exp\left[\frac{f \cdot x_b}{k_B \cdot T \cdot N(t)}\right]$ the Bell's relation for the off rate of a single closed bond. Bell used this approach for a cluster of parallel bonds under constant external force and demonstrated the existence of a critical force above what rupture can happen. Seifert extended it by studying the influence of various parameters such as the loading rate, the rebinding rate and the transducer stiffness. This approach is attractive because it is simple and gives analytical solutions. However, it may not be relevant in some cases such as when we consider the cluster rupture irreversible [71].

Second, we discuss the stochastic approach. This method describes the time evolution of the probability of each possible states. For instance, the probability $p(i)$ that i bonds are closed at time t can be described as follow [71, 72]:

$$\frac{dp_i}{dt} = k_u(i+1) \cdot p_{i+1} + k_r(i-1) \cdot p_{i-1} - \left[k_u(i) + k_r(i)\right] \cdot p_i \tag{3.7}$$

with $k_u(i) = i \cdot k_0 \cdot \exp\left[\frac{f \cdot x_b}{k_B \cdot T \cdot i}\right]$ and $k_r(i) = k_r \cdot (N_0 - i)$ the reverse and forward rates between the possible states i. Using the stochastic approach, Erdman and Schwarz studied the bistability of a cluster under no external force applied [73]. Compared with the deterministic approach, the full stochastic approach is more accurate and relevant for a wider range of cases. However the analytical calculation of the solution is more complicated and thus simulation tools such as Monte Carlo methods may be required.

3.2.2 Multiple bonds in series: implementation of length into Bell model

Here we develop an extension of the Bell model by implementing a length parameter in order to describe the failure of systems with multiple bonds in series. We discuss the relevance of this model later in the computational size effect result section where we test it against simulation results.

We consider a system defined by a series of N identical bonds (same energy barrier E_b and same potential width x_b). We apply a force f on one extremity and fix the other extremity.

First of all, we derive the expression of the system off rate k_S, that is the probability of failure of the system. We define a system failure as the breaking of one bond. Neglecting any other interactions (bending stiffness, etc.), we can assume that each bond is applied the same force f. Assuming that the bonds are independent each other, we use the Bell model description for the off rate k of each bond. Thus we have

$$k = \omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B \cdot T}\right) \tag{3.8}$$

As discussed earlier, the bond off rate k represents the probability of failure of the bond. The probability of failure of the entire system k_S is given by

$$k_S = 1 - k_{S,0} \tag{3.9}$$

where $k_{S,0}$ the probability of zero-failure of the system. Since each bond failure probability k is independent, we have

$$k_{S,0} = (1-k)^n \tag{3.10}$$

Thus

$$k_S = 1 - (1-k)^n \tag{3.11}$$

For system with a long series of bonds, we obtain

$$k_S \approx nk \quad for\, n >> 1 \tag{3.12}$$

Therefore, the probability of failure of a system with a large number of multiple bonds in series is given by

$$k_S \approx n\,\omega_0 \exp\left(-\frac{E_b - f \cdot x_b}{k_B \cdot T}\right) \tag{3.13}$$

This expression describes the simple fact that the more bonds the system has in series the more likely one bond at least among them is supposed to break and thus the more likely the system is supposed to fail. Then, we expect the mean rupture force to decrease with the number of bonds in series, that is the length of the system.

Now, by implementing the pulling rate v into a second expression of the system off rate k_S and identifying both expressions, we can obtain a relation between the rupture force f and pulling rate. Different approaches may be offered, here we discuss one approach.

The system off rate k_S is also defined as the inverse of the system lifetime τ_S . If we assume that it can be described as $\frac{v}{x_b} = \frac{1}{\tau_S} = k_S$ then eq. (3.8) gives us the following rupture force expression

$$f(n,v,E_b,x_b) = -\frac{k_B T}{x_b}\ln(n) + \frac{k_B T}{x_b}\ln(v) - \frac{k_B T}{x_b}\ln(x_b\omega_0) + \frac{E_b}{x_b} = -a\ln(n) + b' \qquad (3.14)$$

Thus, in this model, the rupture force of the system decreases logarithmically with the number of bonds or length of the system.

3.2.3 Coupled strength models

Structures of biological materials can be quite complex and their strength may rely on a variety of physical parameters and phenomena such as biopolymer elasticity and the strength of weak and covalent bonds. In order to obtain an accurate description of the behavior of such materials under mechanical loading, theoretical models may combine different theories. Here, we illustrate this approach through two models that were developed by Rief [74, 75].

First we briefly review a model for biopolymer extensibility [74]. This model combines the Worm Like Chain (WLC) elasticity model with a thermodynamic two-state description extended from Bell's model. The polymer is made of a series of folded protein domains and is stretched at constant speed. The external force is calculated from the WLC model. From the calculated force and the two-state potential profile, Bell's model provides the probability of unfolding of each domain. Then Monte Carlo simulation is used to solve the equations and to provide the force versus extension curve corresponding to a particular protein structure. The

model can be used to measure the parameters of the two-state potential by fitting the simulation curve of force versus loading rate with the experimental results.

Second, we discuss an elastic bond network model for protein unfolding mechanics [76]. This approach combines an elastic model of a network of bonds with irreversible bond fracture kinetics. The network is subjected to external forces. The elastic model considers bonds as identical springs and calculates the force applied on each bond. Through a Bell based model, the unfolding force is predicted. In general, this model may be applied to a variety of different protein structures. It further enables one to study anisotropy and protein unfolding mechanics. Also, it has been suggested that this model can be used with loading rates similar to experimental ones, and thus direct comparisons between experiments and the model are possible [76]. However, it has limitations such as that fracture is irreversible, and that low forces observed in protein rupture cannot be explained.

3.3 Summary of the rupture force calculations based on Bell's model and utilized in our studies

The rupture force can be predicted based on different formulations. For our studies, we will utilize three distinct models and compare them with our mesoscale simulation results. Two of these methods are directly derived from the concept put forth in Bell's model (discussed in Section 3.1.1). The rupture force

$$f_{\text{crit}} = E_b / x_b, \tag{3.15}$$

corresponds to the force to completely diminish the energy barrier, defined as the critical force (see also schematic in Figure 3.1), leading to instantaneous rupture of the bond. For forces smaller than f_{crit} ($f < f_{\text{crit}}$), there exists still a finite probability that the bond breaks. The dynamics of bond rupture for these cases can be estimated based on Bell's model. The following expression is derived from Kramers' theory by Evans and Ritchie, to express the rupture force at a constant loading rate $r = K\,v$ [63], where

$$f(v,E_b,x_b) = \frac{k_B T}{x_b}\ln(v) - \frac{k_B T}{x_b}\ln\left(\frac{k_B T \cdot \omega_0}{K \cdot x_b}\right) + \frac{E_b}{x_b} = a\ln(v) + b\,. \tag{3.16}$$

In eq. (3.16), K is the stiffness of the force transducer and $v = \Delta x / \Delta t$ is the constant pulling speed at which the protein structure is deformed. This relation predicts that the strength depends logarithmically on the pulling velocity. In the following, we refer to this equation as the "Evans model". Another model is given by

$$f(v, E_b, x_b) = \frac{k_B T}{x_b} \ln(v) - \frac{k_B T}{x_b} \ln(x_b \omega_0) + \frac{E_b}{x_b} = a \ln(v) + b. \tag{3.17}$$

This expression is a direct rearrangement of the Bell off rate expression (eq. (3.1)). The assumption made to include the velocity in the expression given by eq. (3.1) is to equal the pulling velocity to the distance to break one bond divided by the lifetime of the bond, that is, set $v = \Delta x / \Delta t = k \cdot x_b$. For further details regarding this model and its derivation we refer the reader to reference [77]. In the following sections, we refer to the model presented in eq. (3.17) as the "Bell model".

Both expressions (eq. (3.16) and (3.17)) predict a logarithmic dependence of the rupture force with respect to the pulling speed, where the parameter a denotes the slope in the f-ln(v) domain, and the parameter b the intercept. Thus the bond energy landscape characteristics x_b and E_b can be determined by fitting one of these equations with the f-ln(v) plot obtained from experiments or simulations. By fitting the slope a, one can obtain the x_b value. Then, by assuming $\omega_0 = 10^{13} s^{-1}$ [78], one can fit the intercept b and obtain the E_b value.

4 Mechanics of alpha-helical protein arrangements

The plan of this chapter is as follows. We begin with a presentation of the computational setup to study the mechanics of AH protein arrangements. This includes a detailed description of the mesoscopic model formulation, the fitting procedure, as well as validation (Sections 4.1 and 4.2). Then we present the applications of the mesoscale model. Section 4.3 is dedicated to a presentation and discussion of a sensitivity study of the computational results on the parameters of the model. Section 4.4 presents the discussion of computational results obtained to study the shear strength of parallel AH protein arrangements. Section 4.5 is dedicated to a presentation and discussion of computational results obtained to study the length dependence of strength properties of AH protein structures in single and parallel arrangements. We conclude in Section 4.6 with a summary, a discussion and an outlook to future research.

4.1 Mesoscale model formulation for alpha-helical protein arrangements

In the following sections we describe our atomistic based multiscale simulation approach used to develop a mesoscale description of alpha-helical protein domains. The setup of the coarse-grained model for alpha-helical protein domains is based on the specific geometry of an alpha-helix, which features a linear array of turns or convolutions; during rupture, any one of these convolutions ruptures as reported in earlier simulation studies [77]. In our model, each convolution is represented by one bead, so that an entire alpha helix is represented by a linear combination of multiple beads that can move independently.

4.1.1 Coarse-graining approach

To achieve the coarse-grained description, the entire sequence of amino acids that makes up the alpha-helical structure is replaced by a collection of beads that are linked to their neighbors by a single type bond (see Figure 4.1(A)). An alpha-helix consists of a series of convolutions, where each of them features 3.6 residues. This protein secondary structure is

stabilized through the presence of H-bonds between the O atom of residue n and the N atom of residue n+4 (there are 3-4 H-bonds between turns in average).

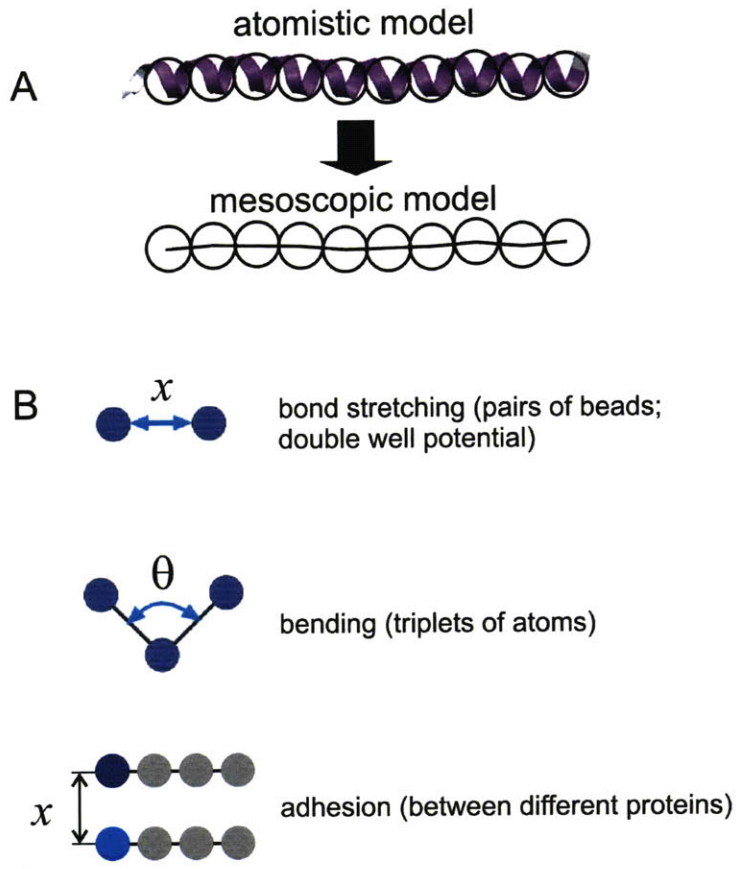

Figure 4.1. Schematic of the coarse-graining procedure, in which we replace the full atomistic representation by a mesoscopic bead model. A pair of beads represents one turn in the alpha-helix (also called a convolution), and thus 3.6 residues (each bead also has the corresponding mass). In the atomistic representation, the folded states of the turns are stabilized by the presence of 3-4 H-bonds between turns. In the mesoscopic bead model, this is represented by using a double-well potential to describe the energy landscape under bond stretching. Different alpha-helix chains interact with a Lennard-Jones potential to describe intermolecular adhesion (the parameters of this potential are adapted to reflect adhesion properties associated with different amino acid sequences).

Our coarse-grained model captures the main structural and energetic features of an alpha-helical protein domain. Earlier full atomistic molecular dynamics simulations have confirmed that H-bonds break in clusters of 3-4 (corresponding to one convolution) for pulling velocities lower than 0.3 m/s [77]. Thus, to capture the role of the convoluted structure on alpha-helical protein strength, the size of one bead (the smallest unit of our system) represents one convolution of the alpha-helix with the corresponding mass that reflects all atoms that are

represented by that bead. Our model is set up to study the alpha-helix protein domain strength for pulling velocities lower than 0.3 m/s.

The beads within an individual alpha-helix interact according to a bond potential and an angle potential (see Figure 4.1(B)). We choose a double-well bond potential in order to capture the existence of two equilibrium states for a convolution, corresponding to the folded and unfolded configuration (see Figure 4.2 for the energy landscape and snapshots of atomistic geometries of the folded and unfolded states).

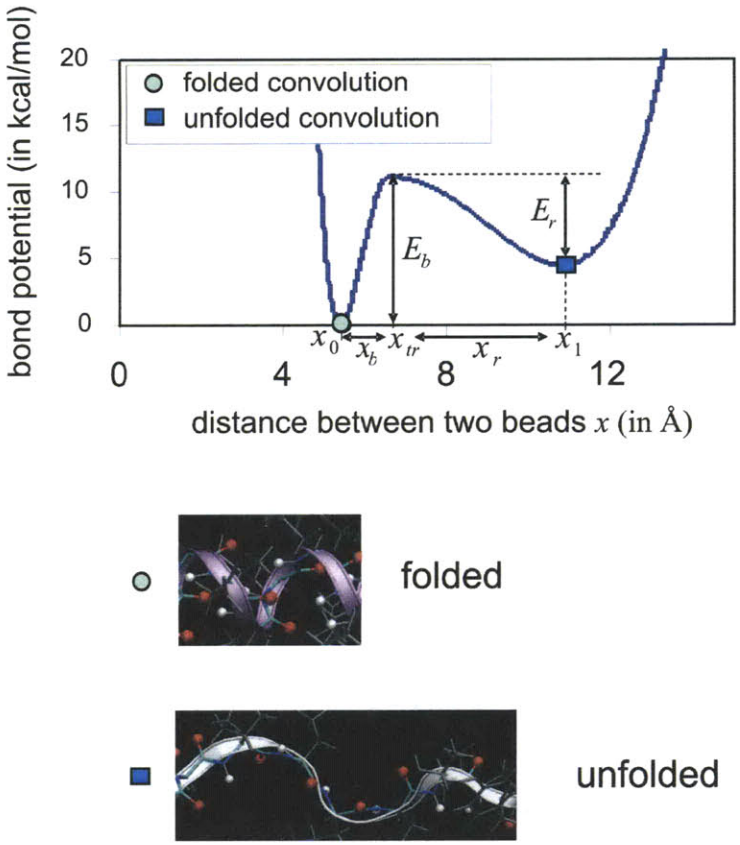

Figure 4.2. Double-well profile of the bond stretching potential in the bead model, representing the energy landscape associated with unfolding of one convolution (this model and the parameters are derived completely from full atomistic simulations [77]). The numerical values of the equilibrium states (x_0 and x_1), energy barriers (E_b and E_r) and the transition state x_{tr} are obtained from geometric analysis of the alpha-helix geometry, as well as full atomistic simulations. The transition state (local energy peak of the potential) corresponds to the breaking of the 3-4 H-bonds between two convolutions of the alpha-helix. After failure of these weak bonds, the convolution unfolds to a second equilibrium state with a larger interbead distance. Under further loading, its covalent bonds begin to be stretched which leads to a second increase of the potential at large deformation. The model does not involve explicit solvent; rather, the effect of solvent on the breaking dynamics of alpha-helical convolutions is implicitly captured in the effective double-well potential. Through this formulation, the bond potential can describe the microscopic details of the rupture mechanism of the 3-4 H-bonds between each convolution under force, and the transition from the folded states to the unfolded states of convolutions through an energy barrier that separates the two states. Yet, the description is sufficiently "coarse" so that it enables a significant computational speedup and efficiency compared with the full atomistic description. To characterize interactions between alpha-helical protein domains in a parallel arrangement we use a Lennard-Jones (also abbreviated as LJ) potential between pairs of beads of different strands. This describes the adhesion energy between alpha-helices (see Figure 4.1(B), lower part).

The mathematical expression for the total energy of the system is given by

$$E = E_T + E_B + E_S, \tag{4.1}$$

where E_T is the total tensile energy, E_B the total bending energy and E_S the total intermolecular interaction energy. The total tensile energy is given by the sum over all pair-wise interactions,

$$E_T = \sum_{pairs} \phi_T(x) \cdot \tag{4.2}$$

The double-well bond potential $\phi_T(x)$ is given by (see also Figure 4.2 for a schematic):

$$\phi_T(x) = \begin{cases} \frac{E_b}{x_b^4}(x - x_{tr})^2 \cdot (x - x_{tr} - \sqrt{2} \cdot x_b) \cdot (x - x_{tr} + \sqrt{2} \cdot x_b) + E_b, & x < x_{tr} \\ \frac{E_r}{x_r^4}(x - x_{tr})^2 \cdot (x - x_{tr} - \sqrt{2} \cdot x_r) \cdot (x - x_{tr} + \sqrt{2} \cdot x_r) + E_b, & x_{tr} \le x \end{cases} . \tag{4.3}$$

The first equilibrium with reaction coordinate x_0 (first potential minimum) corresponds to the folded state of one turn of an alpha-helix under no force. The transition state (energy barrier E_b), with position x_{tr} (peak of the potential between two wells), corresponds to the breaking of the 3-4 H-bonds between two turns of the alpha-helix. After failure of these weak bonds, the turn unfolds to a second equilibrium state. This corresponds to the second potential minimum with a larger interbead distance, x_1. Under further loading, its backbone bonds begin to be stretched which leads to a second increase of the potential. The parameters x_b and E_b represent the distance and energy barrier required to unfold one convolution. Similarly, x_r and E_r correspond to the refolding process.

The total bending energy is given by the sum over all triples of beads,

$$E_B = \sum_{triplets} \phi_B(x) \cdot \tag{4.4}$$

The angular potential is given by

$$\phi_B(\theta) = \frac{1}{2} K_B(\theta - \theta_0)^2, \tag{4.5}$$

where K_B relates to the bending stiffness of the molecule EI, θ as the interbead angle (in triplets of atoms), and θ_0 as the equilibrium angle. In order to distinguish the bending stiffnesses of a folded alpha-helix and an unfolded alpha-helix (which entails a severe structural change of the protein), we use a stiffness parameter K_B that depends on a reaction coordinate x such as:

$$K_B = K_{B,fold}\left[(1-\alpha)\frac{1}{\pi}\left(-\arctan\left(100(x-x_{tr})\right)+\frac{\pi}{2}\right)+\alpha\right] \tag{4.6}$$

$$\text{with } \alpha = \frac{K_{B,unfold}}{K_{B,fold}}, \tag{4.7}$$

the ratio between the bending stiffness parameters of the unfolded state $K_{B,unfold}$ and folded state $K_{B,fold}$ for one convolution. These two bending stiffness parameters are given by

$$K_{B,fold(unfold)} = \frac{3EI_{fold(unfold)}}{x_{0(1)}} \tag{4.8}$$

with $x_{0(1)}$ denoting the equilibrium bead distance which corresponds to the equilibrium distance of one folded (unfolded) convolution, and $EI_{fold(unfold)}$ as the bending stiffness of a folded (unfolded) alpha-helix. This formulation based on a reaction coordinate is similar in spirit as reactive force field formulations based on chemical bond orders (such as Tersoff, ReaxFF, and others, as reviewed for example in reference [79]).

The total intermolecular interaction energy E_S is given by the sum over all pair-wise interactions between beads of different alpha-helical protein domains,

$$E_S = \sum_{pairs} \phi_S(x) \cdot \tag{4.9}$$

The adhesion potential ϕ_S is described by a Lennard-Jones potential,

$$\phi_S(x) = 4\cdot\varepsilon\cdot\left[\left(\frac{\sigma}{x}\right)^{12}-\left(\frac{\sigma}{x}\right)^{6}\right]. \tag{4.10}$$

The energy minimum ε of the Lennard-Jones potential corresponds to the adhesion energy per convolution between two alpha-helices in equilibrium divided by 3.51 (this factor takes

into account the fact that we do not use any cut-off distance for the Lennard-Jones potential and thus a bead interacts with more than one bead in the other protein structure, therefore the adhesion energy for each bead is the same as the adhesion energy estimated from full atomistic simulation); and x is the distance between mesoscale particles. The zero-crossing distance σ is linked to the equilibrium bead distance d_0 (equilibrium spacing between two alpha-helices) by

$$d_0 = \sqrt[6]{2} \cdot \sigma. \tag{4.11}$$

4.1.2 Parameter fitting: linking atomistic and mesoscale models

The parameters in the coarse-grained formulation are determined through a fitting procedure against geometric properties of alpha-helices as well as full atomistic molecular dynamics simulation results in explicit solvent. We fit the energy barrier measured from MD simulation to the energy barrier in the mesoscale model formulation. The mass of each bead corresponds to the approximate average mass of one turn or 3.6 residues, leading to 400 amu.

The parameters of the tensile double-well potential are introduced in eq. (4.3) (see also Figure 4.2). We find x_0 = 5.4 Å for the equilibrium bead distance of the folded state, which corresponds to the length of one folded convolution. The distance x_b between the folded state equilibrium and the transition state corresponds to the distance to break 3-4 H-bonds which leads to the unfolding of the convolution, and the parameter E_b is the corresponding energy barrier (full atomistic molecular dynamics simulations [77] have confirmed that H-bonds in convolutions indeed break in clusters). These two parameters are determined from fitting against full atomistic simulations of tensile loading experiments of alpha-helical domains, for a range of pulling rates below 0.3 m/s [77]. We find x_b = 1.2 Å and E_b = 11.1 kcal/mol by fitting a theoretical strength model based on the Bell model to these full atomistic studies. The equilibrium bead distance of the unfolded state, x_1, is determined by fitting the mesoscopic force-strain curve against the atomistic simulation results at large deformation. The parameter describes at what strain levels a alpha-helix convolution is completely unfolded, and when further strain leads to significant stiffening due to stretching of the protein backbone. Therefore we adjust x_1 to so that the angular point between the plateau regime and the backbone stretching regime occurs at the same strain as in the atomistic simulation curve. We

find x_1= 10.8 Å, which corresponds to twice the length of the folded state. Lastly, the energy barrier E_r to refold a "broken" convolution must be smaller than E_b, since the folded state is the most favorable state for a convolution in equilibrium. Based on a suggestion put forth in [80], we determine $E_r = 0.6\ E_b$ (it is noted, however, that the resulting mechanical properties of the alpha-helix is insensitive to variations of choices of E_r, as long as $E_r < E_b$ as we have confirmed by direct simulation [results not shown here]). A quantitative comparison of direct simulations of alpha-helical protein domains under mechanical loading has shown that the coarse-grained model accurately represents the rupture properties and agrees quantitatively with full atomistic results (for further details, see [81]).

The parameters of the angle potential are introduced in eq. (4.5) and eq. (4.6). The value of the equilibrium angle θ_0 is 180 degrees, based on the overall molecular geometry of the alpha-helical structure. The bending stiffness parameters $K_{B,fold}$ and $K_{B,unfold}$ are linked to molecular parameters as described by eq. (4.8) and therefore can be determined from full atomistic simulations or experiments of bending studies of alpha-helical protein domains. For the folded configuration, we use the results reported in references [82] and [79]. We find $K_{B,fold}$= 21.589 kcal/mol/rad^2, which corresponds to a persistence length of approximately 6.5 nm. For the unfolded configuration, we use the typical persistence length of a free polypeptide chain, which is about 0.4 nm [83]. Thus we find $K_{B,unfold}$ = 0.665 kcal/mol/rad^2.

The two parameters of the intermolecular potential are introduced in eq. (4.10) and eq. (4.11). The parameter ε is linked to the adhesion energy per convolution between two alpha-helices in equilibrium, and the parameter d_0 to the equilibrium spacing between two alpha-helices. They are determined from full atomistic simulations as shown in Figure 4.3 for the reference case. The full atomistic simulation results are obtained in collaboration with Zhao Qin.

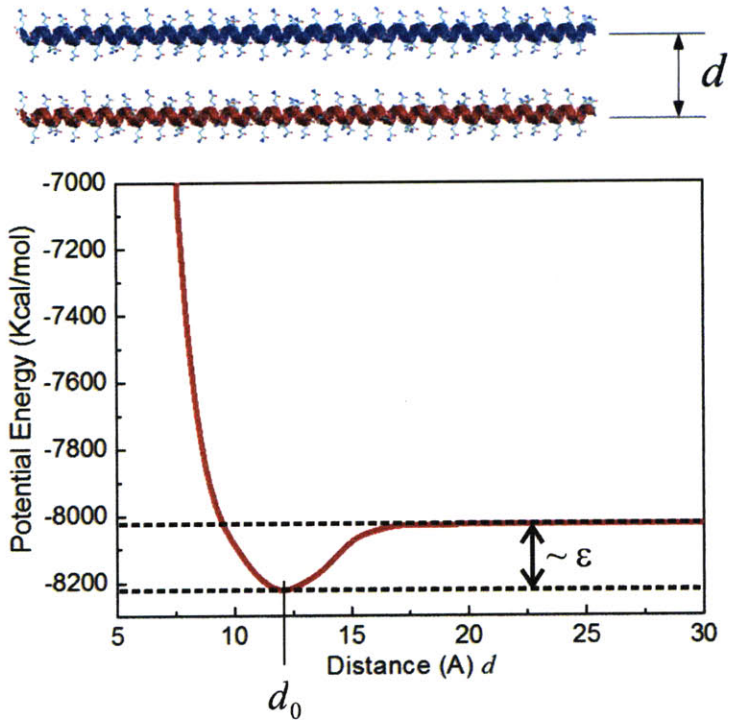

Figure 4.3. Full atomistic model with implicit solvent calculation with CHARMM, used here to identify the intermolecular adhesion between two parallel alpha-helices. This simulation reveals that the adhesion energy is g=1.262 kcal/mol/Å, that is 6.815 kcal/mol per convolution; with an equilibrium spacing of 12.1 Ångstrom. The full atomistic simulation is carried out for an alpha-helix sequence made of solely GLN amino-acids. The full atomistic simulation results are obtained in collaboration with Zhao Qin.

For the reference case we consider two identical alpha-helices in parallel. Each alpha-helix contains 100 glutamine (abbreviated by the symbol GLN) amino acids, with an approximate length of 150 Å. The two alpha-helices are kept parallel to each other at a distance of d by fixing all the C_a atoms in each chain while all other atoms (including side chains) are free to move. We utilize the EEF/CHARMM-19 force field with excluded volume implicit solvation model in our atomistic simulations [84, 85]. We consider a series of molecular arrangements with systematically increasing intermolecular distance d; for each case we record the potential energy after energy minimization. Figure 4.3 shows the resulting curve, enabling us to extract adhesive properties between pairs of two alpha-helices (where we systematically change the intermolecular spacing between two parallel alpha-helices as the controlled parameter). The molecular dynamics studies reveal that the adhesion energy is γ=1.262 kcal/mol/Å (corresponding to the minimum of the adhesive energy function), reached at d_0 =12.1 Å equilibrium spacing, leading to a Lennard-Jones distance parameter σ=10.8 Å. The adhesion

energy per convolution is linked to the Lennard-Jones parameter ε, which is determined to be $\varepsilon = 6.815$ kcal/mol.

The complete set of parameters of the reference mesoscopic model (for the GLN sequence) and their physical meaning is summarized in Table 1.

Model parameters	**Numerical values**
Equilibrium bead distance of the folded state x_0 (in Å)	5.4
Equilibrium bead distance of the unfolded state x_1 (in Å)	10.8
Distance between folded state and transition state x_b (in Å)	1.2
LJ distance parameter between beads of two AHs, σ (in Å)	10.8
Energy barrier between folded state and transition state E_b (in kcal/mol)	11.1
Energy barrier between unfolded state and transition state E_r (in kcal/mol)	6.7
Energy minimum between beads of two AHs, ε (in kcal/mol)	6.815
Equilibrium angle θ_0 (in degrees)	180
Bending stiffness parameter of the folded state $K_{B,fold}$ (in kcal/mol/rad^2)	21.589
Bending stiffness parameter of the unfolded state $K_{B,unfold}$ (in kcal/mol/rad^2)	0.665
Mass of each mesoscale bead (in amu)	400

Table 4.1. Summary of parameters of the mesoscale model, derived from geometrical analyses and atomistic simulations, corresponding to eq. (4.1-4.12), as well as the discussion presented throughout the parameter fitting section (for the reference GLN sequence).

4.2 Validation of the mesoscale model

Before studying any applications of the mesoscale model of AH protein structures, the first step is to validate our model. To complete this goal, here we validate the mesoscale results against full atomistic simulation results for two different loading experiments: tensile loading and shear loading experiments. Since with our model we aim to study deformation of AH

based material under tensile or shear stress, we limit the validation step to these two loading tests.

4.2.1 Tensile loading experiments

Figure 4.4 depicts the entire force-strain curve for a stretching experiment on the 14 bead mesoscopic model of an alpha-helix with a length of 70.2 Angstroms, at a temperature of 300 K and a pulling rate of 0.1 m/sec. The upper part represents a schematic of this tensile loading experiment carried out on both the coarse-grained and full atomistic models. In the lower part, the force-strain curve of the mesoscale model shows the three typical regimes observed in full atomistic simulations: an elasticity regime at low strain, an energy dissipation regime which corresponds to the unfolding of the 13 bonds (13 peaks on the curve), and the subsequent regime of stretching of the backbone bonds. The peaks of force, which correspond to the AH rupture force, fit atomistic results very closely. Both full atomistic and mesoscale model predict an initial rupture force, referred to as the first force peak (FFP) or angular point (AP), of approximately 350 pN, with a slight increase as the strain is increased.

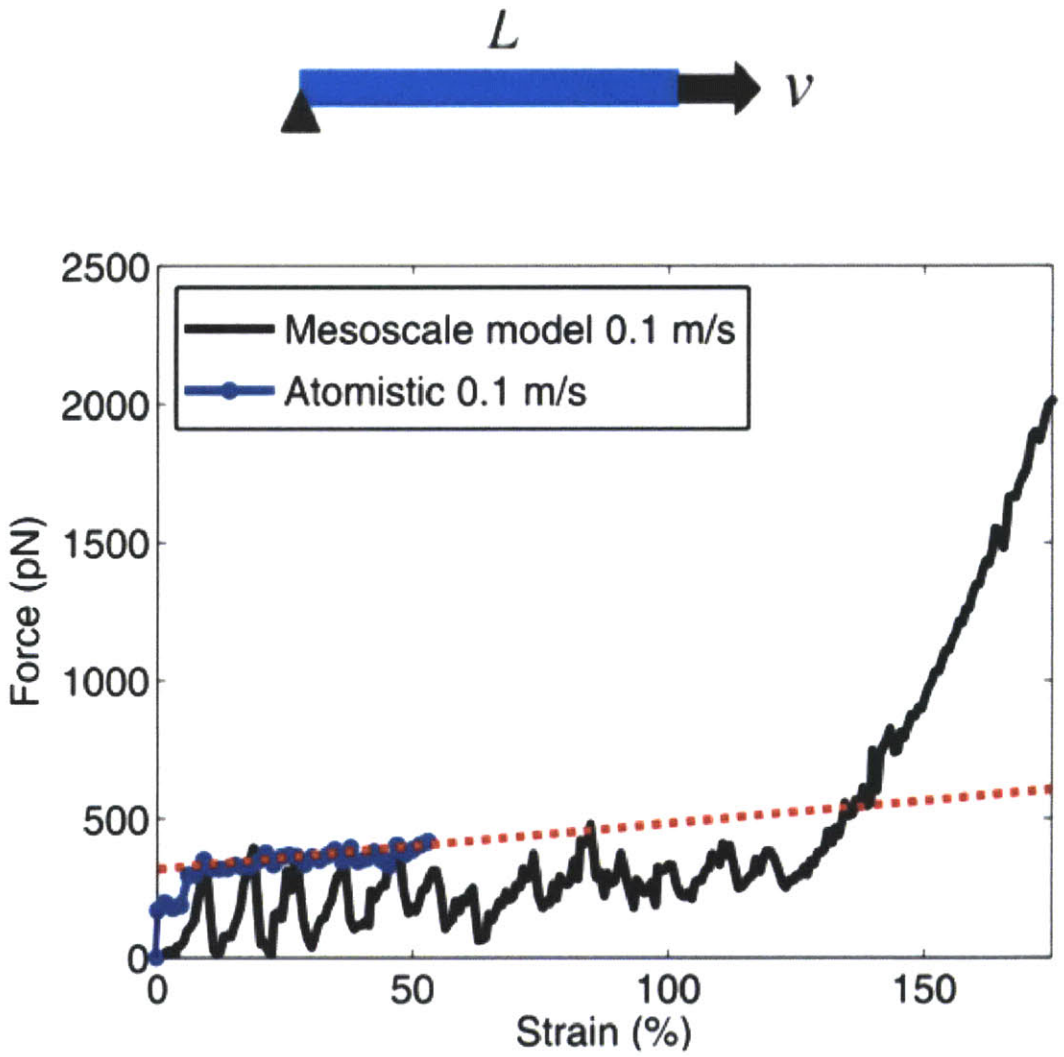

Figure 4.4. Entire force-strain curve for a stretching experiment on the 14 beads mesoscopic model of an alpha-helix (with a length of 70.2 Angstroms at a temperature of 300 K and a rate of 0.1 m/s). The upper part represents a schematic of this tensile loading experiment carried out on both the coarse-granied and full atomistic models. In the lower part, the mesoscale curve shows the three typical regimes observed in full atomistic simulations: an elasticity regime (before the first peak occurs), an energy dissipation regime which corresponds to the unfolding of the 13 bonds (corresponding to the 13 peaks on the curve), and the regime of stretching of the backbone bonds. The peaks in the mesoscale model, corresponding to the rupture forces, agree very well with the rupture force measured from full atomistic simulation. We note that the full atomistic simulation curve presents only small fluctuations since the atomistic system have about three thousand times more particles (atoms) than the mesoscale model, and thus a larger number of interactions contribute to the damping of force relaxation during unfolding of convolutions. The dotted line approximates the increase of rupture force with increasing strain. In agreement with atomistic simulations reported earlier [82], the third regime sets in at approximately 135% strain.

Figure 4.5 presents the validation of our mesoscale model for tensile loading experiments by direct comparison of the strength of alpha-helical protein domains with full atomistic results

of the rupture mechanics of an individual alpha-helix protein. The plot shows the rate dependence of the unfolding force for both models. The straight line in this plot corresponds to the predictions by the theoretical Bell model (eq. (3.17)) fit to the atomistic results, discussed in Section 3.3. The mesoscale model is in very good agreement with the full atomistic simulations, validating the fitting of the mesoscopic bond potential.

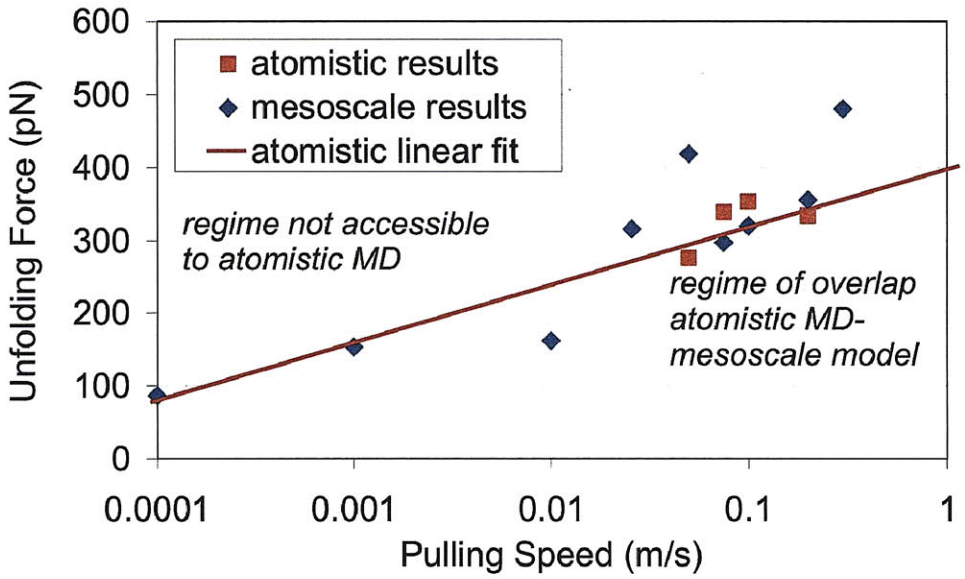

Figure 4.5. Validation of the mesoscopic model for tensile loading experiments by comparison with full atomistic results (MD results taken from reference [77]). The plot shows the rate dependence of the unfolding force for both models. The mesoscale model is in very good agreement with the full atomistic simulations, validating the fitting of the mesoscopic bond potential. The Figure further illustrates that the mesoscale model is capable of reaching much slower pulling rates than those accessible to full atomistic simulation studies, here shown for the slowest pulling speed of 0.0001 m/sec.

The results shown in Figure 4.5 illustrate a key advantage of the coarse-grained model in reaching much longer timescales than what could be achieved in full atomistic simulations (the current limit in MD simulations is 0.01 m/sec, whereas we have easily reached a 100 fold increase in accessible timescales using our mesoscale model). The model is capable of reaching time-scales of several microseconds and longer with a quantitative accuracy comparable with full atomistic MD simulations. Such relatively long simulations can be carried out within several days of computational time (on a single Intel Xeon CPU). In comparison, MD simulations of the dynamical behavior at fractions of microseconds can take weeks and months of computational time (even on a large parallelized simulation setup). This reflects a considerable speedup due to the coarse-graining approach, while the model is still capable of describing the small- and large-deformation force-strain response characteristics

(e.g. softening at ≈10% strain and stiffening at ≈135% strain) as well as strength values quite accurately.

Reaching long timescales is a key advantage for computational models because they get closer to experimental conditions and thus the computational results can be compared to experimental data more easily than traditional atomistic models. For instance, experimental results of stretching and breaking single AH domains [77] (with a length of less than 100 Å) report forces between 140 and 240 pN during unfolding, corresponding to the force level predictions at ultra-slow pulling speeds. Opposed to it, the full atomistic model is limited to shorter timescales (pulling rate above 0.01 m/s) and thus comparing or discussing results that are pulling rate dependent with experimental data is a difficult task. For instance, one can use theoretical models such as Bell model to link computational data to experimental data over the timescale gap but we will see in Section 4.3.2 that this approach has some limitations. We refer the reader to Section 4.3.2 for a more detailed discussion and analysis of timescale issues.

Two different definitions of the unfolding force (or strength, or rupture force) of an individual AH. Depending on the studies, we may use a different definition of the unfolding force (referred also to strength or rupture force) of an AH protein. Overall, we use two different definitions of the unfolding force. In the figure above (Figure 4.5), we refer to "unfolding force" as the "first (force) peak" (FFP) or Angular Point (AP) of the force strain curve of a stretching experiment. Later, we may refer to "unfolding force" as the average unfolding force during the entire unfolding regime. We call it the mean unfolding force (MUF). In some studies, we prefer this method against measuring only the first peak, because it averages among all bond rupture events and avoids strong sensitivity to the resolution of the data. However, this method is more time consuming and sometimes even not relevant (e.g. for the size-effect study). As a consequence, if the mesoscale is used to fit closely to full atomistic simulation results, we suggest to use an energy barrier slightly higher than the one from the reference model (13 kcal/mol instead of 11.1 kcal/mol) in order to compensate the decrease of the measured unfolding force due to force relaxations after the peaks, Figure 4.6 shows the results of the validation of the mesoscopic model using this time the MUF definition for tensile loading experiments by direct comparison of the strength of alpha-helical protein domains with full atomistic results of the rupture mechanics of an individual alpha-helix protein. Note that we use an energy barrier slightly higher than the one from the

reference model (13 kcal/mol instead of 11.1 kcal/mol) in order to compensate the decrease of the measured unfolding force due to force relaxations after the peaks

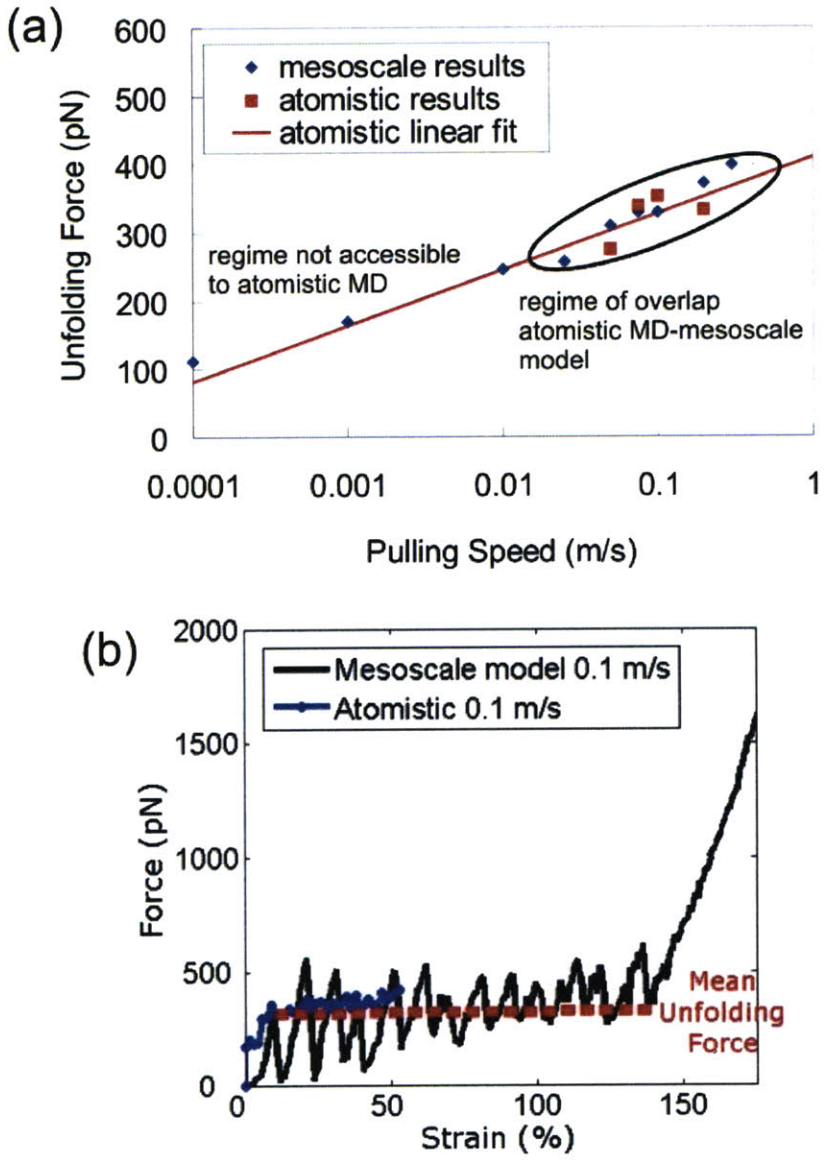

Figure 4.6. Validation of the mesoscopic model using the mean unfolding force definition (MFU) for tensile loading experiments by comparison with full atomistic results (MD results taken from reference [86]). Subplot (a) shows the rate dependence of the unfolding force for both models. Subplot (b) depicts the entire force-strain curve for a stretching experiment on the 14 beads mesoscopic model of an alpha-helix.

The mean AH unfolding force, which corresponds to the mean force of the energy dissipation regime, fit atomistic results rather closely. Both full atomistic and mesoscale model simulations predict a rupture force of approximately 350 pN.

4.2.2 Shearing experiments

Figure 4.7 (upper part) shows a schematic of the coarse-grained model setup to validate the mesoscale model for describing shearing experiments. We consider two aligned alpha-helical proteins. One strand has its end fixed and the second strand is pulled in the opposite direction at a constant velocity v =0.1 m/s. The coordinates of each bead as well as the force applied at the pulled extremity are stored every ten picoseconds. Then the deformation mechanisms can be observed using Visual Molecular Dynamics (VMD) [77]. We plot the force against strain (or equivalently, the extension) with a bin resolution of two bins per Angstrom. The shear strength is defined as the average force in the plateau regime observed in a typical force strain curve of a shearing experiment. Figure 4.7 lower part presents the entire force strain curves for a shearing experiment carried out using the mesoscopic model and the full atomistic model of two parallel alpha-helices (we use 18 GLN residues per alpha-helix strand in the full atomistic model and thus five beads per alpha-helix strand in the mesoscopic model, a pulling velocity of 0.1 m/s and a temperature of 300 K). The full atomistic model is initially fully equilibrated with CHARMM in EEF implicit solvent and is pulled by using the CHARMM Steered Molecular Dynamics (SMD) module with a pulling velocity of 0.1 m/s. The full atomistic simulation results are obtained in collaboration with Zhao Qin.

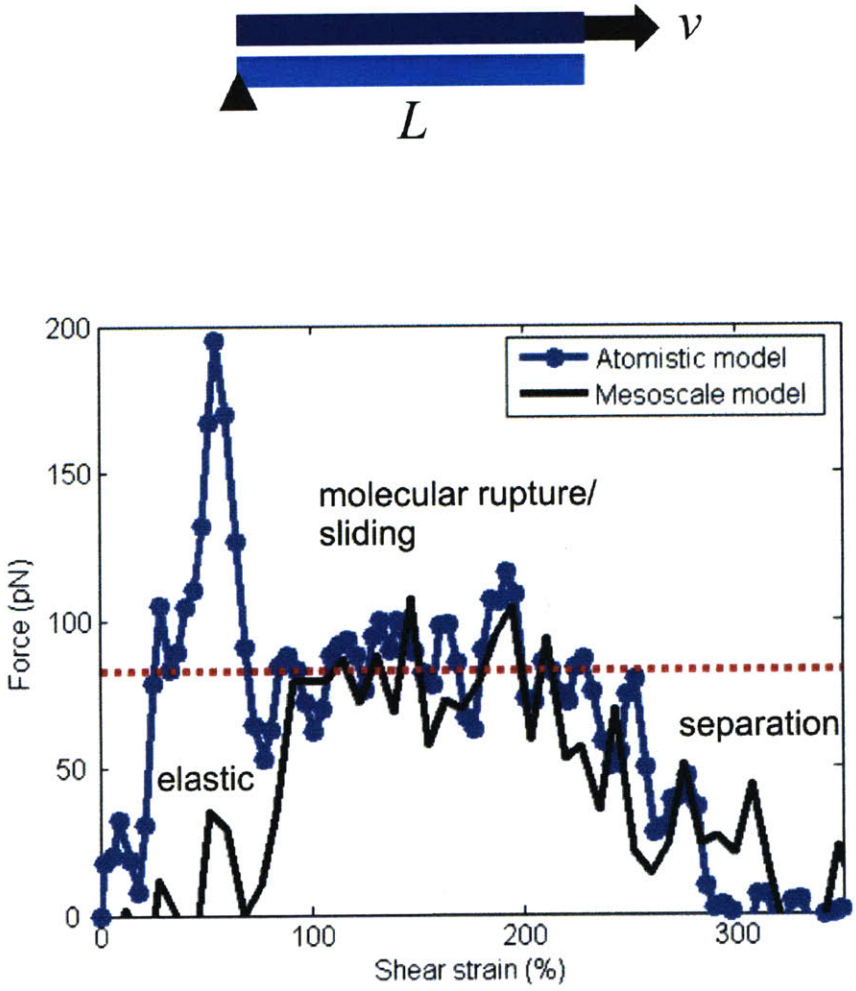

Figure 4.7. Shear strength of an assembly of two AHs. The upper part shows a schematic of the coarse-grained model setup for the shearing of two aligned AH proteins, focusing on a simple assembly of two protein constituents. This model represents a simple representation of more complex biological structures, but enables us to carry out a systematic engineering analysis of how deformation is accommodated in these protein structures. One strand has its end fixed and the second strand is pulled in the opposite direction at a constant velocity v =0.1 m/s. The lower part depicts the entire force strain curves for a shearing experiment on the mesoscopic model and the full atomistic model of two parallel alpha-helices (we use 18 GLN residues per alpha-helix strand in the full atomistic model and thus 5 beads per alpha-helix strand in the mesoscopic model, a pulling velocity of 0.1 m/s and a temperature of 300 K). The shear strain is defined as the ratio between the extension and the equilibrium distance between two alpha-helix strands (12.1 Å). Both curves show three different regimes. First, an elasticity regime as we start shearing, then a plateau regime, which corresponds to the sliding of the pulled strand along the fixed strand and lastly a zero force regime once the two strands separate. The average of the force value over the plateau regime gives the mean shear strength (indicated with the horizontal red dashed line for the mesoscale model). The difference between the full atomistic and mesoscopic model at the first peak region is that the full atomistic model requires initial unfolding of some convolutions before sliding as shown in Figure 4.8. The full atomistic simulation results are obtained in collaboration with Zhao Qin.

The shear strain is defined as the ratio between the extension and the equilibrium distance between two AH strands (12.1 Å). Both curves show three different regimes. First, an elastic regime as we start shearing, then a plateau regime, which corresponds to the sliding of the pulled strand along the fixed strand and lastly a zero force regime once the two strands separate. The average of the force value over the plateau regime gives the mean shear strength (indicated with the horizontal red dashed line for the mesoscale model). The plateau regimes of both models are very similar, except at the beginning of the plateau where the atomistic model presents a high initial force peak. This initial force peak reaches almost 200 pN, whereas the mean shear strength of the entire plateau is 90 pN, and therefore very close to the one from the mesoscale model (83 pN). The reason for this initial force peak could be that the shearing for the full atomistic model requires initial unfolding of some convolutions (generated by the first high force peak) before sliding of the two alpha-helices, or perhaps equilibration effects at the beginning of the simulation. It is also possible that in the full atomistic case the system has a much larger number of particles (atoms), and thus many more interactions (not only within backbones but also the side chains), and therefore a significantly larger number of degrees of freedom. These differences may explain the more distinct force peak seen in the full atomistic simulations. Thus this result illustrates the limitations of the mesoscale model in that it can not resolve some atomistic-level details. This could be addressed in future work.

Figure 4.8 shows the dynamics of the shearing of two alpha-helices by depicting several snapshots; for both the mesoscale and full atomistic models. The full atomistic simulation results are obtained in collaboration with Zhao Qin. As before, we use 18 GLN residues per alpha-helix strand in the full atomistic model and thus five beads per alpha-helix strand in the mesoscopic model, a pulling velocity of 0.1 m/s and a temperature of 300K. Both models show that the predominant deformation mechanism is shearing; with some unfolding convolutions appearing at the extremities in the full atomistic case. The unfolding of these convolutions appeared at the beginning of the shearing experiment and that could explain the high initial force peak of the shear force observed only in the case of the atomistic model.

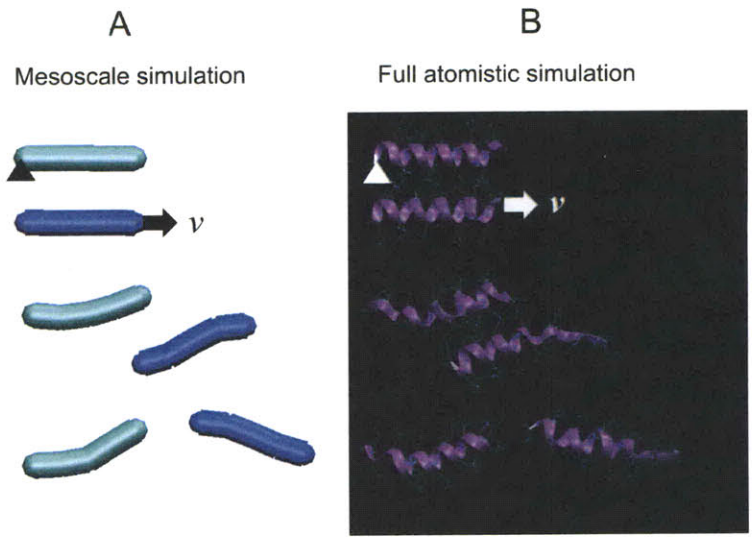

Figure 4.8. Dynamics of the shearing of two alpha-helices; for the mesoscale (panel A) and the full atomistic model (panel B). We use 18 GLN residues per alpha-helix strand in the full atomistic model and thus 5 beads per AH strand in the mesoscopic model, a pulling velocity of 0.1 m/s and a temperature of 300K. Both models show that the predominant deformation mechanism is shearing; with some unfolding convolutions appearing at the outermost parts of the alpha-helix in the full atomistic case (this is not captured in the coarse-grained model). The full atomistic simulation results are obtained in collaboration with Zhao Qin.

4.3 Sensitivity study

We have developed and calibrated (by fitting against full-atomistic MD results) a mesoscale model of AH protein arrangements, representing one convolution as a pair of mesoscale bead particles (Figure 4.2). This represents a reference system, which agrees well with the predictions from Bell's or Evans' models (Figures 4.4 and 4.5). In this section, we proceed with a systematic variation of the parameters in the mesoscale model and study their impacts on the computational results.

4.3.1 Energy landscape effects

We now proceed with a systematic variation of the parameters in the mesoscale model, including: E_b, x_b, E_r, x_r as well as the bead mass M. We vary these parameters one by one, while fixing the others at their standard values as defined in the reference model (see Table 4.1 and the fitting parameter section, Section 4.1.2).

4.3.1.1 On the rate dependence of the rupture force

4.3.1.1.1 Influence of E_b and x_b

Figure 4.9 shows plots of the unfolding force versus pulling speed obtained with the double-well potential mesoscale model for different values of x_b and E_b. The range of pulling speeds is taken from 0.01 m/s to 0.3 m/s, which corresponds to the atomistic simulation regime where 3-4 H-bonds (that is, one whole convolution) break simultaneously [77].

Figure 4.9 (a) shows the plots of the unfolding force versus pulling speed obtained for different x_b values, varying from 0.5 Å to 2 Å. The other potential parameters remain at their standard values as in the reference model. According to eq. (3.16) and eq. (3.17), changes in the parameter x_b lead to a change in slope and a change in the intercept with the y-axis. This behavior is observed in Figure 4.9 (a). Figure 4.9 (b) shows the plots of the unfolding force versus pulling speed obtained for different E_b values, varying from 5 kcal/mol to 20 kcal/mol. As before, the other potential parameters remain at their standard values as in the reference model. According to eq. (3.15), changes in the parameter E_b lead to a change in the intersection value, while the slope is maintained. This behavior is observed in Figure 4.9 (b), albeit this prediction fails for relatively small values of the energy barrier (≤ 7 kcal/mol). For the sake of clarity, we do not show the plots for high energy barriers (E_b values between 20 and 100 kcal/mol), which correspond to much larger rupture forces. The linear fits of these two plots show the well-known logarithmic rate dependence of the unfolding force. Overall, the variations of slopes and intercepts between these linear fits show that E_b and x_b have an influence on the logarithmic rate dependence. We will now measure their influence carefully and quantitatively compare them with the predictions of the Bell model.

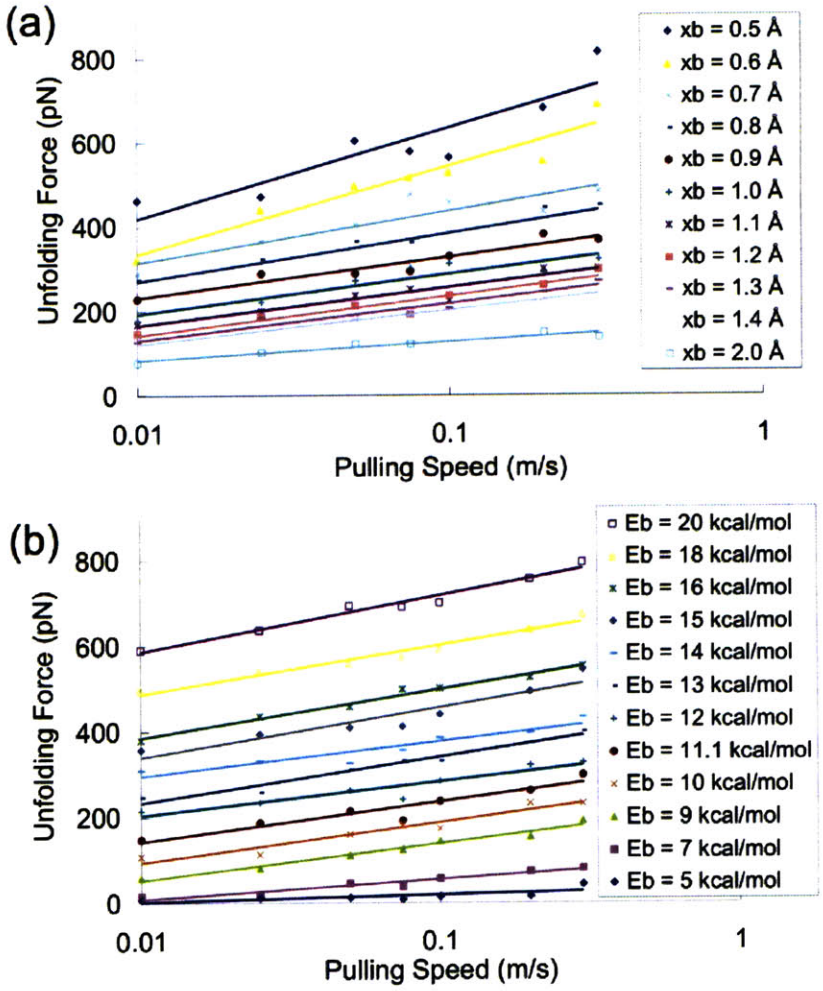

Figure 4.9. Plots of the unfolding force versus pulling speed obtained with the double-well potential mesoscale model for different values of x_b and E_b (range of pulling speeds is taken from 0.01 m/s to 0.3 m/s which corresponds to the atomistic simulation regime where 3.6 H-bonds (one whole convolution) break simultaneously [80, 87, 88]. Subplot (a) shows the plots of the unfolding force versus pulling speed obtained for different x_b values varying from 0.5 Å to 2 Å, while the other potential parameters remain at their standard values. Subplot (b) shows the plots of the unfolding force versus pulling speed obtained for different E_b values varying from 5 kcal/mol to 20 kcal/mol (that is, between 8 k_BT and 34 k_BT), while the other potential parameters remain at their standard values. The linear fits of these two plots show the well-known logarithmic rate dependence of the unfolding force. The variations of slopes and intercepts between these linear fits show that both x_b and E_b have an influence on the logarithmic rate dependence.

Figure 4.10 depicts the plots of the slope a and the intercept b of the logarithmic rate dependence of the unfolding force, for variations of E_b and x_b,

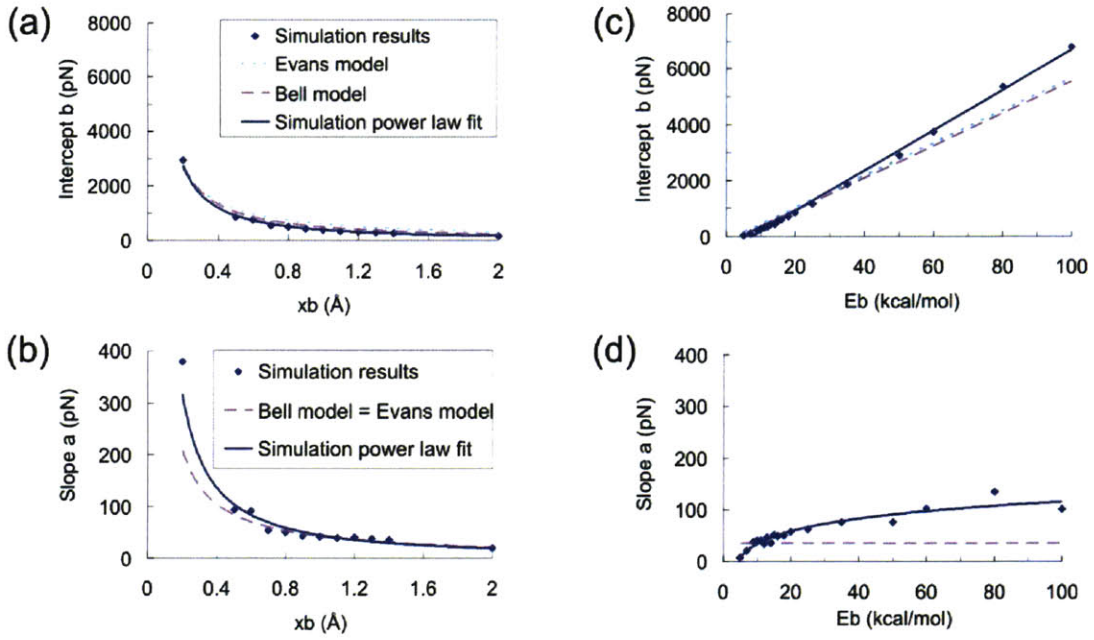

Figure 4.10. Plots of the slope a and the intercept b of the logarithmic rate dependence of the unfolding force, versus x_b and E_b, obtained with the mesoscale model, the Bell model and the Evans model. Subplots (a-b) show the dependence of the slope a and intercept b on x_b, respectively. Subplots (c-d) show the dependence of the slope a and intercept b on E_b, respectively.

The graphs show results obtained with the mesoscale model (diamonds for the data and solid line for the fit which is either linear, power law or logarithmic), the Bell model (medium-dashed line) and the Evans model (short-dashed line). For the slope plots, the Evans model is not represented because it is identical to Bell model. Figure 4.10(a) plots the slope a as a function of x_b, and shows a close agreement between simulations and Bell model (note that for this data, E_b= 11.1 kcal/mol). In the given range of x_b, the maximum error of the predicted value (according to the Bell model) against the measured x_b value (from the slope) is below 25%, and tends to maximize as x_b reaches rather small values (corresponding to very shallow energy barriers). Figure 4.10(b) plots the intercept b as a function of x_b. The results reveal a close agreement between simulations, Bell and Evans models to describe the influence of x_b on the intercept b (note that as before for this data, E_b= 11.1 kcal/mol). In the given range of x_b for a fixed value of E_b= 11.1 kcal/mol, the maximum error of the Bell

model to predict the measured x_b value from the intercept b is below 20% and is found at larger values of x_b. For the Evans model the maximum error is below 40%.

Figure 4.10(c) plots the slope a as a function of E_b (note that for this data x_b=1.2 Å fixed). The results show that the Bell model fails to describe the influence of E_b on the slope as observed in the simulations. Most importantly, the Bell model predicts no dependence of the slope a on E_b, whereas the simulation data clearly reveals a dependence of a on E_b. This dependence could be fit empirically by the following logarithmic relation,

$$a = \left(35.48 \ln(E_b / (\text{kcal/mol})) - 48.31\right)\text{pN}, \qquad (4.12)$$

The disagreement between the Bell model prediction and the measurement from simulation is largest for very small values of E_b and for very large values of E_b. Within the range of energy barriers between 9-18 kcal/mol (that is, for a range of 15-30 k_bT), which corresponds to the range of weak bonds (e.g. H-bonds and clusters of H-bonds), the slope a does not vary significantly since the variation is smaller than 45%. If one uses the Bell model to determine x_b from the measurement of the slope a, then, given that x_b is inversely proportional to the slope a, the error of x_b is equal to the error on the slope a, which leads to an error on x_b smaller than 45% as well (thus comparable to the error observed in the previous plots). This result suggests that in the range of weak bond energies, the Bell model is a suitable model to describe the energy landscape. We note that many applications of the Bell model in the literature are indeed within the range of weak bond energies [89]. For E_b values lower than 9 kcal/mol, we find a strong decrease of the slope when E_b decreases. A possible reason is that the critical velocity below which the Bell model predicts a negative force value enters the studied range of velocities for values below E_b=7 kcal/mol (E_b=5 kcal/mol for the Evans model). Thus, in this range of velocities and for low E_b values, we would expect a deviation from the Bell model prediction (for instance through a smooth transition between the Bell linear fit to a very low rupture force in an asymptotic regime [68, 90, 91]).

In light of Evans' derivation, this analysis may also suggest that the assumption made in Kramers' theory that the energy barrier is much greater than the thermal energy breaks down for E_b below 9 kcal/mol in this studied range of velocities. The disagreement at very large values of E_b could be explained by the fact that for values larger than 20 kcal/mol, the Bell

model for "weak bonds" is no longer applicable, and other protein strength models suitable for strong bonds must be used. Figure 4.10(d) plots the intercept b as a function of E_b (note that as before for this data x_b=1.2 Å fixed). The comparison reveals a close agreement between our simulations and both the Bell and Evans models. Within the E_b value ranges corresponding to weak bonds, the highest relative errors for the Bell and the Evans models are below 15% and 30%, respectively. For larger E_b values, the highest error is below 15% for both Bell and Evans models.

4.3.1.1.2 Influence of E_r and x_r

Figure 4.10 shows the plots of the slope a and the intercept b for variations of potential parameters that relate to the second local unfolded equilibrium (see Figure 4.2 for a schematic and corresponding atomistic geometry as well as a description of parameters used to characterize this part of the energy landscape). These are parameters x_r and E_r (corresponding to the transition distance and energy barrier for refolding of a convolution). It is noted that these two parameters are not included in Bell's or Evans' models, respectively. The parameter x_r varies between $0.1\,x_b$ and $4\,x_b$. The parameter E_r varies within a range between 0 and E_b, so that the first equilibrium remains the most probable state (that is, a global minimum).

Figure 4.10(a) plots the influence of variations of x_r on the slope a, and shows that x_r does not have a significant influence on the slope. This is evident from the fact that the linear fit to the simulation data is almost horizontal. Figure 4.10(b) plots the influence of variations of x_r on the intercept b. The results suggest that x_r does not have a significant influence on the intercept, except for values lower than x_b. For instance, the intercept is approximately 100% higher for x_r=$0.1\,x_b$ than for x_r=$2\,x_b$ (the plot shows a fit of a power law to the data). The results suggest that the Bell model may not be relevant to describe an energy landscape where the width of the second potential equilibrium is shorter than the width of the first potential equilibrium. Figure 4.10(c) and Figure 4.10(d) show that E_r does not have any significant influence on either the slope a or the intercept b. In conclusion, the parameters of the second

unfolded equilibrium do not affect the logarithmic rate dependence of the unfolding force, provided that x_r is higher than x_b. Conversely, if x_r is significantly lower than x_b, the Bell model is not relevant.

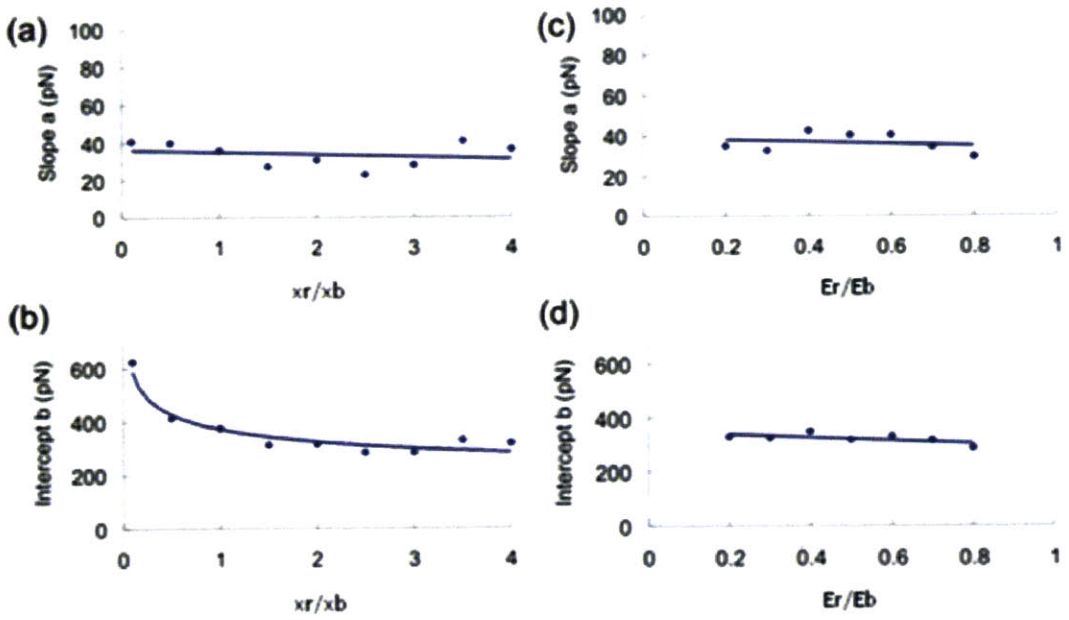

Figure 4.11. Plots of the slope a and the intercept b of the logarithmic rate dependence of the unfolding force, versus x_r and E_r (properties associated with the second local unfolded equilibrium), obtained with the mesoscale model (diamonds for the data and solid line for the fit which is either linear, power law or logarithmic). Subplots (a-b) show the dependence of the slope a and intercept b on x_r, respectively. Subplots (c-d) show the dependence of the slope a and intercept b on E_r, respectively. The plots show that the parameters of the second potential well do not significantly influence the rate dependence of the unfolding force except for x_r values below x_b.

4.3.1.2 On the rupture force

4.3.1.2.1 Influence of E_b and x_b

We proceed with an analysis of the influence of E_b and x_b on the rupture force at a pulling speed of 0.1 m/s. Figures 4.12(a-b) show the dependence of the unfolding force on E_b and x_b. As predicted by both Bell's and Evans' models, we observe a linear force dependence on E_b and a relative linear force dependence on $1/x_b$. The critical force expression f_{crit} given by eq.(3.15) qualitatively agree with the scaling behavior found in the simulations. We note that

the critical force typically yields an overestimation of the rupture force compared with simulations and the other models, except for very large energy barriers. This is due to the fact that this expression does not take into account that the probability of the bond to break for a force below the critical force is not zero.

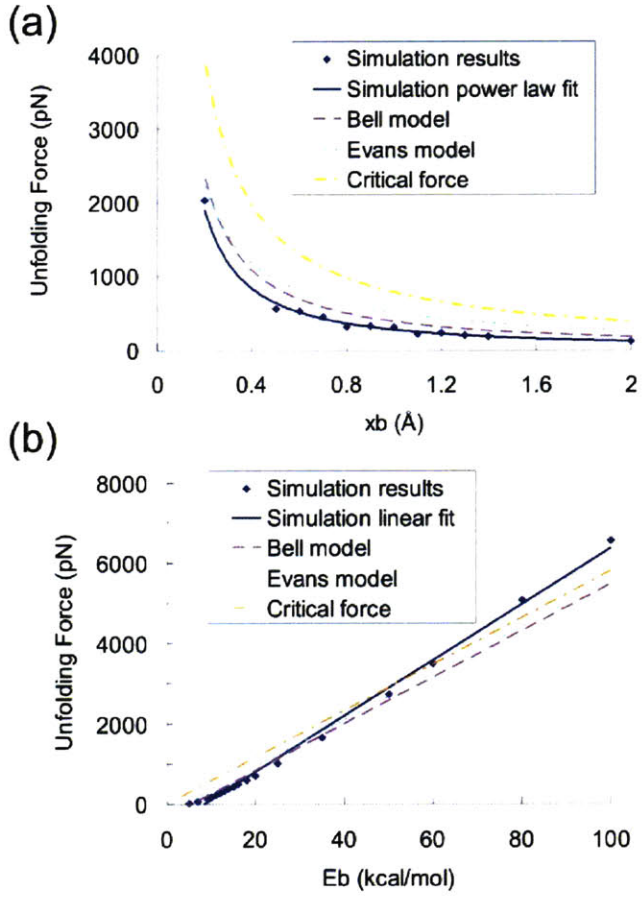

Figure 4.12. Dependence of the unfolding force on x_b and E_b obtained with the mesoscale model, the Bell model, the Evans model and the critical force expression at a pulling speed of 0.1 m/s. As predicted by Bell's and Evans' models as well as the simple critical force expression, we observe a linear force dependence on E_b (subplot (b)) and a relative linear force dependence on $1/x_b$ (subplot (a)).

4.3.1.2.2 Influence of E_r and x_r

We proceed with an analysis of the influence of x_r and E_r on the rupture force at a given pulling speed of 0.1 m/s (as before, since these parameters are not included in the Bell model there is no influence of these parameters predicted from the Bell model). We focus on two aspects, first the value of the rupture force at the first peak in the force-strain plot (see Figure 4.4), and second on the average unfolding force during the entire unfolding regime (as the strain increases until stiffening occurs due to stretching of the protein backbone). Figures 4.13 (a-b) show the influence of x_r and E_r on the first peak in the force-strain plot. The data clearly shows that these parameters do not influence the first force peak.

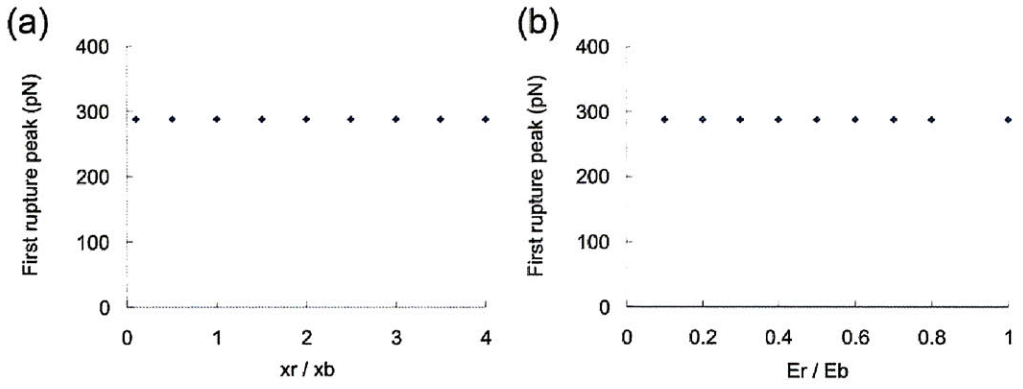

Figure 4.13. Influence of x_r and E_r on the first rupture peak. Subplots (a) and (b) show that the parameters of the second equilibrium do not influence the first bond rupture event (first force peak).

Figure 4.14 depict the influences of x_r and E_r on the average unfolding force. Figure 4.14(a) shows that E_r does not significantly influence the unfolding force. In contrast, Figure 4.14(b) shows that x_r influences the unfolding force for x_r values below x_b.

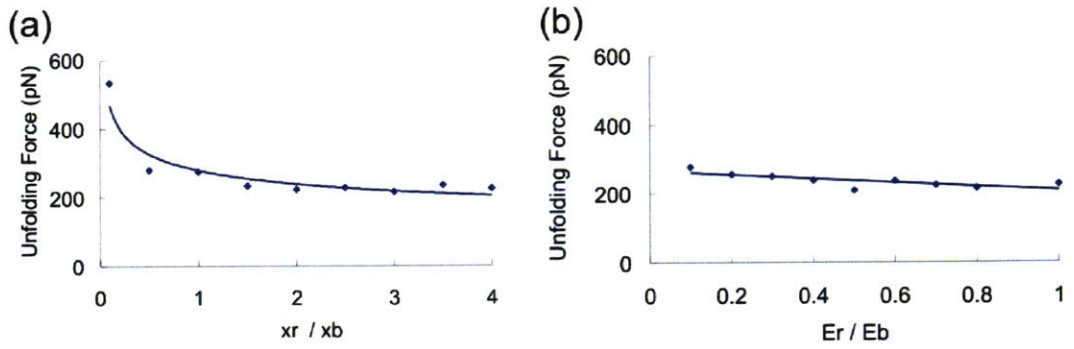

Figure 4.14. Influence of x_r and E_r on the average unfolding force. Subplot (a) shows that E_r does not significantly influence the unfolding force. Subplot (b) shows that x_r influences the unfolding force more strongly for x_r values below x_b.

4.3.1.3 Discussion and conclusion

We have carried out a systematic analysis of the effects of energy landscape parameters on the strength properties of AH protein domains, and compared with commonly used protein strength models. The two most important contributions of this section are:

- The reference system provided us with the starting point for a systematic variation of all relevant parameters, to identify how strength properties depend on the details of the energy landscape and under which conditions the Bell model fails. By systematically varying the energy landscape parameters, we have shown that the Bell model is a reasonable approximation to predict the mechanical strength properties as long as the model parameters are in a certain range: E_b between 9-18 kcal/mol (weak bond energy barrier) and x_r value above x_b value. (see, e.g. Figure 4.10). Conversely, the Bell model breaks down when the model parameters are not in this range. We have also analyzed the dependence of the unfolding force on E_b and x_b (Figure 4.12) and confirmed the predictions from the Bell model.

- We have shown that the properties of the second, unfolded state of a AH convolution do not strongly influence the strength properties, provided that x_r, the distance between the unfolded state and the transition state, is higher than x_b, the distance between the folded

state and the transition state.(see, e.g. Figure 4.11). We have also shown that the energetic properties of the second unfolded state do not influence the rupture strength.

4.3.2 Multi-timescale analysis

An advantage of coarse-grained models compared to full atomistic model is the possibility to study systems over a larger range of timescales since simulations run faster. Here we use our AH mesoscale model to study the rupture force of individual AHs over a wide range of pulling rates. The rate dependence of the rupture force of single molecules have been studied for decades (see review of protein strength model in Chapter 1 as well as reference [89]), but this question remains challenging and still generates numerous discussions based on various approaches or models.

One example of the significance of this question is the actual gap that exists between experimental data and simulation data on rupture force of systems. Understanding and explaining the rate dependence of the rupture force could enable to finally link simulation results to experimental results.

4.3.2.1 Results

Figure 4.15 depicts the rupture force versus pulling rate of the mesoscale model of an individual AH over seven orders of magnitude. We plot the two rupture force definitions (force at first peak and mean unfolding force, see Section 4.2 for description) and compare the curves with the Bell model predictions. Note that we decrease the energy barrier of the tensile double-well potential to 7.04 kcal/mol in order to move the critical rate (that is the rate above what the Bell model rupture force is positive) to accessible values and observe the force vs rate behavior for rates below the critical rate.

Moreover, in order to get closer to experimental conditions, we use a low transducer spring constant, K=0.1 kcal/mol/A^2 (see Section 4.3.3 for a detailed analysis of transducer spring constant effect). The lower plot is a zoom of the upper plot. We observe two regimes for the force at first peak (FFP) and three regimes for the mean unfolding force (MUF). The Bell model prediction line is consistent with the FFP results and the MUF for pulling rates above

the critical rate, so for rates above 0.01 m/s. However, for very high pulling rates (above 100 m/s), the Bell model does not fit well any more the MUF curve that is increasing exponentially. For pulling rates below the critical rate, both MUF and FFP curves converge to a respective asymptotic positive strength.

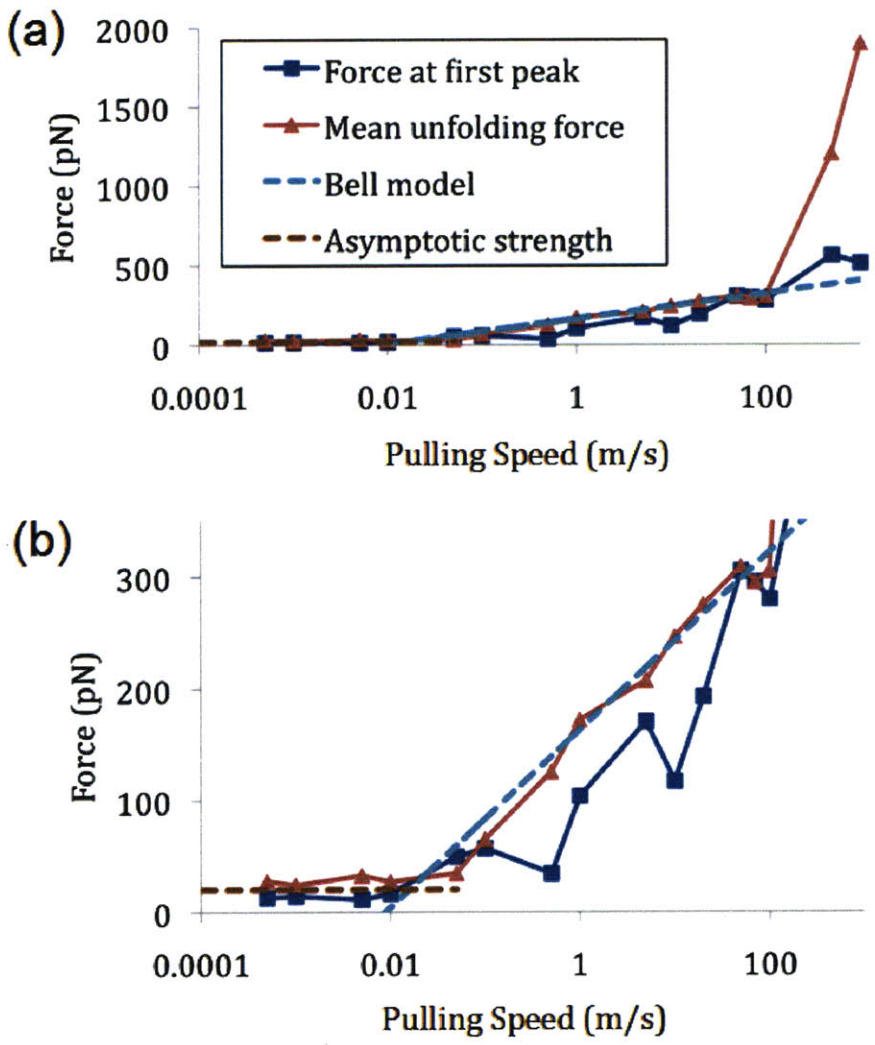

Figure 4.15 Rupture force versus pulling rate of the mesoscale model of an individual AH over seven orders of magnitude. We plot the two rupture force definitions (force at first peak and mean unfolding force) and compare the curves with the Bell model predictions. The lower plot is a zoom of the upper plot. We observe two regimes for the force at first peak (FFP) and three regimes for the mean unfolding force (MUF). The Bell model prediction line is consistent with the FFP results and the MUF for pulling rates above the critical velocity (that is the rate above what the Bell model rupture force is positive), so for rates above 0.01 m/s. However, for very high pulling rates (above 100 m/s), the Bell model does not fit any more the MUF curve that is increasing exponentially. For pulling rates below the critical rate, both MUF and FFP curves converge to a respective asymptotic positive strength. Our results seem to agree qualitatively with the theoretical multi-timescale model for AH developed in [89] in terms of the existence of the asymptotic positive regime followed by the first Bell model regime whose energy barrier value corresponds to the one used to fit the energy barrier of our double-well bond potential.

Figure 4.16 illustrates the rate dependence of the first force peak and of the mean unfolding force (plateau regime). It depicts the force strain curves of the mesoscale model of an

individual AH for different pulling rates v within a range of six orders of magnitude (from 0.0005 m/s to 50 m/s). Both the FFP and MUF increase as the pulling rate increases.

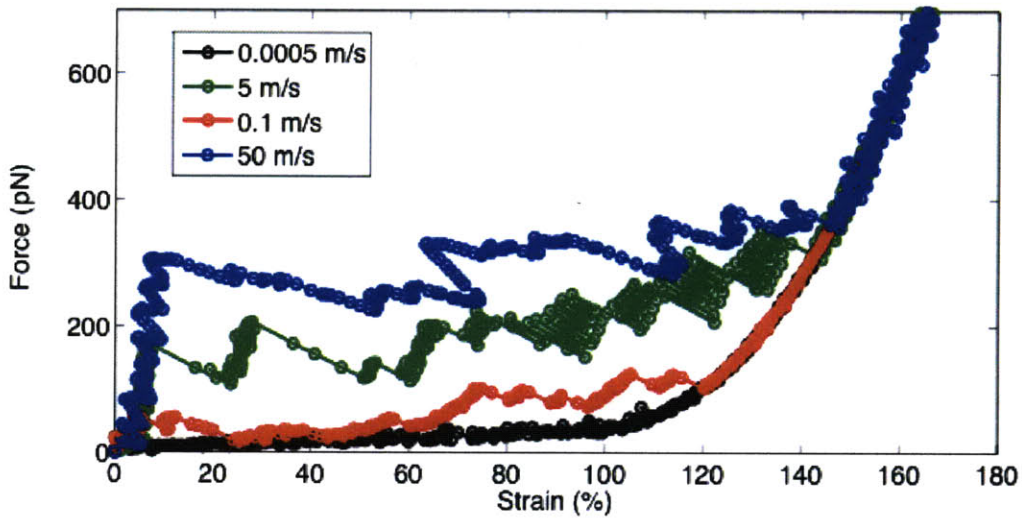

Figure 4.16. Force strain curves of the mesoscale model of an individual AH for different pulling rates v within a range of six orders of magnitude (from 0.0005 m/s to 50 m/s). All plots have the same resolution (2 bins per angstrom). The Figure illustrates the rate dependence of the first force peak and of the mean unfolding force (plateau regime). Both increase as the pulling rate increases.

4.3.2.2 Discussion and conclusions

Our results presented in Figure 4.15 seem to agree qualitatively with the theoretical multi-timescale model for AH protein domains developped in [68, 90, 91] in terms of the existence of the asymptotic positive strength regime (for pulling rates below the critical rate) followed, as the pulling rate increases, by the Bell model regime whose energy barrier value corresponds to the one used to fit the energy barrier of our double-well bond potential. In the first chapter of the Thesis we review several theoretical approaches on protein strength such as Bell's model and Evans'model. These last two models predict a logarithmic rate dependence of the rupture force. Our study tests the validity of these predictions for our double-well potential based coarse-grained model and shows some limitations of such simple models (for pulling rates below the critical rate and for very high pulling rates). Other more complex theoretical models have been developed in order to better explain the dynamic force spectroscopy of single molecules and the rate dependence of their rupture forces [87]. Our results seem also to qualitatively agree with Friddle's model and Hummer's model on the

point that the rupture force vs loading/pulling rate should converge to the Bell model slope at low pulling rates (but still above the critical rate).

Note that experiments are within the following range of loading rates 0.1pN/s-0.1μN/s [92], that corresponds to the range of pulling speeds 1pm/s-0.001m/s (using $v = Kr$, for transducer spring constant values within the range 0.001 kcal/mol/A^2-0.1 kcal/mol/A^2 [83]. Our coarse-grained model can reach pulling speeds as low as 0.0001 m/s with a computational time of a few days. Therefore the mesoscale model is able to reach some experimental timescales.

Given that the energy barrier of the double-well bond potential has been fit using the Bell model on atomistic data within a range of 0.01 m/s and 0.3 m/s (see Section 4.1.2), we may wonder whether the coarse-grained AH model is still relevant for lower pulling rates. First of all, as shown in [89], the deformation mechanisms should remain identical at lower rates. Indeed, it has been theoretically shown that H-bonds are likely to break by clusters of 3-4 H-bonds at low pulling rates. Therefore the hypothesis that an individual AH unfolds by convolution of 3-4 H-bonds is still relevant at low pulling rates. Lastly, as discussed earlier, the coarse-grained AH model seems to describe as well the transition between the Bell model and the asymptotic rupture force regime observed at very low pulling rates as demonstrated in [93]. Therefore the mesoscale AH model may be relevant for any pulling rates below 0.3 m/s.

4.3.3 Influence of the spring constant

Here we study the transducer stiffness dependence of the rupture force of an individual AH. A transducer is a device where one extremity is linked to the single molecule and the other is undergoing a force or displacement. Thus it is used to apply a force or pulling rate to a molecule. A transducer is characterized by a stiffness or spring constant. Since each experiments or simulations may use a transducer of different spring constants, we may wonder what is the influence of the spring constant on the rupture force results. Better understanding the spring constant impact on the rupture force could help one to reduce the gap between results of different methods such as between an experiment and a simulation. The stiffer the molecule sample is the stiffer we want the transducer to be. Thus for simulations we tend to use higher spring constants (e.g. 10 kcal/mol/A^2 is the value we use in our previous studies) than in experiments (between 0.001-0.1 kcal/mol/A^2) since simulations give higher rupture force (in part due to higher pulling rates than in experiments).

Figure 4.17 depicts the rupture force versus transducer stiffness of the mesoscale model of an individual AH over six orders of magnitudes, pulling at a velocity of 0.1 m/s. We plot the two rupture force definitions (force at first peak, FFP, and mean unfolding force, MFU). According the FFP fit, the rupture force increases logarithmically with the transducer spring constant whereas the MFU curve shows a very low spring constant dependence of the rupture force. The circle area underlines the range of transducer spring constants accessible in experiments.

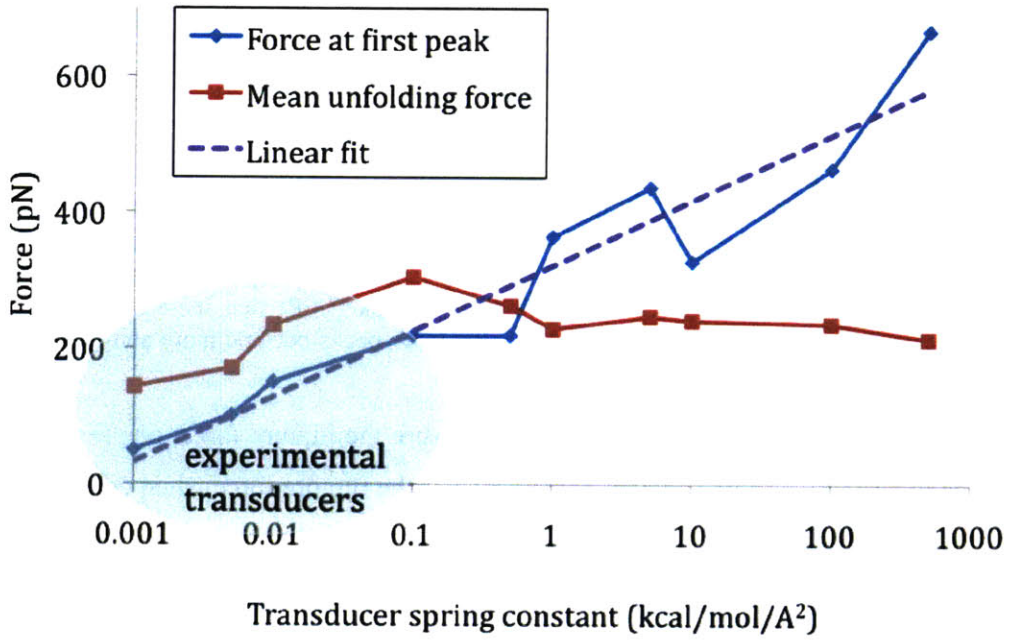

Figure 4.17. Rupture force versus transducer stiffness of the mesoscale model of an individual AH over six orders of magnitudes, pulling at a velocity of 0.1 m/s. We plot the two rupture force definitions (force at first peak, FFP, and mean unfolding force, MFU). According the FFP fit, the rupture force increases logarithmically with the transducer spring constant whereas the MFU curve shows a very low spring constant dependence of the rupture force. The circle area underlines the range of transducer spring constants accessible in experiments.

Figure 4.18 shows the force strain curves of the mesoscale model of an individual AH for different transducer spring constants K within a range of six orders of magnitude (from 0.005 kcal/mol/A^2 to 1 kcal/mol/A^2). All plots have the same resolution (2 bins per angstrom). The figure illustrates the transducer spring constant dependence of the first force peak (FFP) that increases as the spring constant increases. As the spring constant increases, the force peaks become more and more apparent and sharp and the FFP increases.

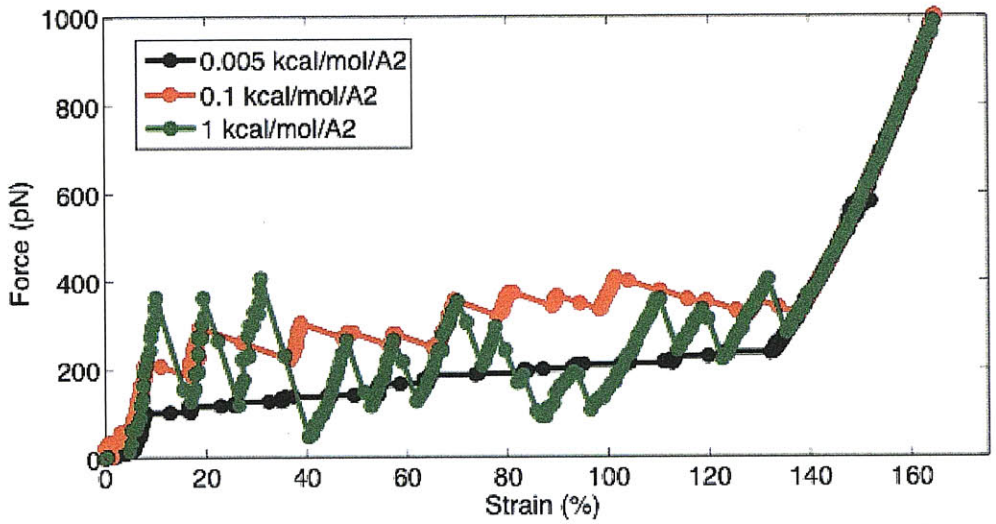

Figure 4.18. Force strain curves of the mesoscale model of an individual AH for different transducer spring constants K within a range of four orders of magnitude (from 0.005 kcal/mol/A^2 to 1 kcal/mol/A^2). All plots have the same resolution (2 bins per angstrom). The figure illustrates the transducer spring constant dependence of the first force peak (FFP) that increases as the spring constant increases. As the spring constant increases, the force peaks become more and more apparent and sharp and the FFP increases.

To conclude, if one is using the FFP method to measure the rupture force, one may take into account the logarithmic spring constant dependence of the rupture force when comparing with results from other methods.

4.3.4 Influence of bending stiffness

In the method section, we present the coarse-grained development of the individual AH mesoscale model. Once a bond break, the bending stiffness of the new unfold segment decreased a lot in order to take into account the change of bending stiffness between an alpha-helical structure stabilized by HBs and a unfold polypeptide chain. Here we investigate how this change of bending stiffness may impact the force strain curve.

Figure 4.19 shows the force strain curves for AH models without (black curve) and with (red curve) change of bending stiffness when a convolution unfolds, pulling at a velocity v=0.001 m/s. We observe that the model with change of bending stiffness is stronger than the model without this implementation as soon as we start pulling.

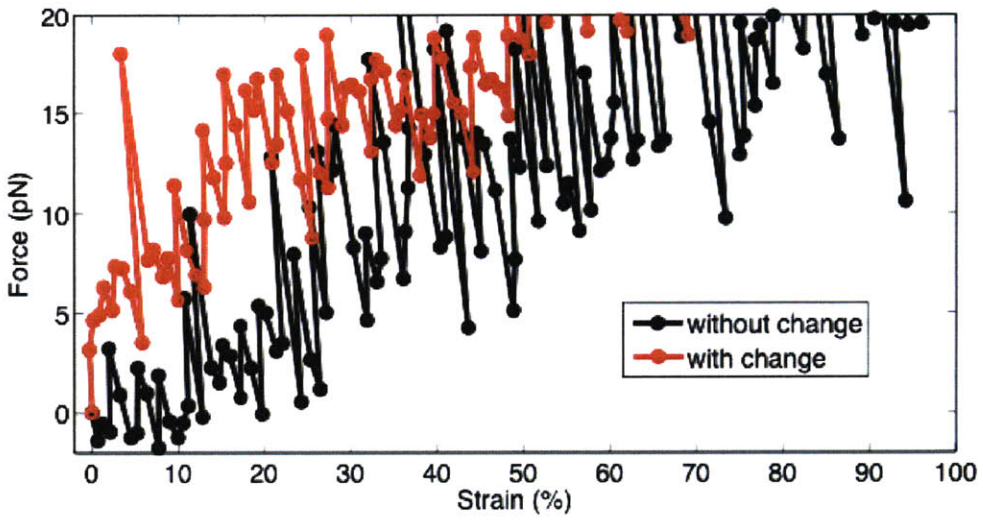

Figure 4.19. Force strain curves for AH models without (black curve) and with (red curve) change of bending stiffness when a convolution unfolds, pulling at a velocity v=0.001 m/s. We observe that the model with change of bending stiffness is stronger than the model without this implementation as soon as we start pulling.

4.3.5 Influence of the bead mass on the rupture force

Figure 4.20 shows the bead mass influence on the average unfolding force at a pulling speed of 0.1 m/s. The results illustrate a very slight dependence on the rupture force. This analysis is, however, hypothetical since the mass of each bead is not an adjustable parameter (the mass is given by the number of atoms associated with each "bead").

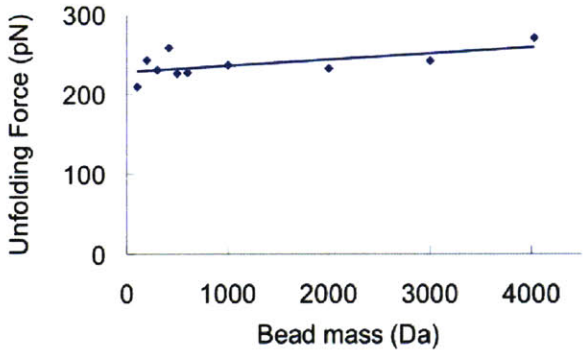

Figure 4.20. Influence of the bead mass on the rupture force. The continuous line shows a linear fit to the data obtained from mesoscale simulations. The plot shows that the bead mass influences the rupture force only marginally.

4.4 Shear strength: Amino acid sequence dependence of nanoscale deformation mechanisms in alpha-helical protein filaments

We proceed as follows. First we report a series of full atomistic studies to analyze the dependence of intermolecular adhesion on the amino acid sequence. Then we perform a series of coarse-grained simulations to test a broad range of intermolecular adhesion values and how they influence the resulting deformation behavior in alpha-helical protein filaments.

Amino acid sequence	**Adhesion energy per length (kcal/mol/Å)**	**Adhesion energy per length (pN)**	**Equilibrium distance (Å)**
ALA-ALA...ALA	0.126	8.7	9.5
GLU-GLU...GLU	0.23	16.0	12.2
VAL-VAL...VAL	0.305	21.2	11.1
Human vimentin 2B segment	0.367	25.5	12.5
TRP-TRP...TRP	1.26	87.5	11.1
GLN-GLN...GLN	1.262	87.6	12.1
Carbon nanotubes	1.59	110.5	11.14

Table 4.2. Characteristic parameters of the adhesion energy between two identical AHs for different amino acid sequences and also two carbon nanotubes (with a radius of 4 Å). The adhesion energy and the equilibrium distance of parallel AHs are obtained from the full atomistic model described in Figure 4 and in the fitting parameter section, Section 4.1.2. The adhesion energies are given per length in kcal/mol/Å and also in pN. From our theoretical model (eq. (4.13)), we expect the shear strength between two alpha-helices is equal to the adhesion energy per length. The full atomistic simulation results are obtained in collaboration with Zhao Qin.

4.4.1 Interprotein adhesion from full atomistic simulation

We first determine the characteristic parameters of the adhesion energy between two identical alpha-helices for different amino acid sequences, following the CHARMM simulation protocol outlined above (see fitting parameter section, Section 4.1.2), but here adapted to include a variety of amino acid sequences (other than GLN as used for the reference case). The results are presented in Table 4.2 and compared with the ones of some known engineering materials such as two carbon nanotubes [93] (each one with a radius of 4 Å).

Within the list of tested amino acid sequences, the range of values of the adhesion energy per length goes from 8 to 90 pN. We observe that the adhesion energy per length may vary by a factor ten depending on the amino acids. This significant variation of adhesion energy between amino acids is due to the difference of properties of their side chains. For instance, the alanine, the valine and the glutamic acid have low adhesion energies probably because their side chains are relatively small (for ALA and VAL), which means smaller surface of interaction between two chains, or charged (GLU), which induces stronger repulsion between two identical alpha-helices. On the contrary, tryptophan and glutamine based sequences have relatively high adhesion energies probably because their side chains are polar (thus more attracted to each other) and relatively large (leading to a larger surface of interaction). The equilibrium distance varies between 9 and 12.5 Å. Comparing alpha-helix intermolecular adhesion energy results with the case of engineering materials such as carbon nanotubes, we find that the two carbon nanotubes we considered in earlier studies [81, 89] have slightly higher adhesion energies of 110 pN.

4.4.2 Effects of varying amino acid sequences on shearing of alpha-helical protein filaments

Now we use the mesoscale model to study the effect of the adhesion energy on shearing of two parallel alpha-helices. This system is interesting because it represents a subunit of a typical alpha-helical protein filament such as an intermediate filament (see Figure 1.2). We start from the deformation of the two GLN based protein domains, take their adhesion energy as a reference, and change the adhesion energy parameter to lower and higher values as

discussed above. We then compare the behavior of the system for different adhesion values and discuss the results in light of the adhesion energies reported above.

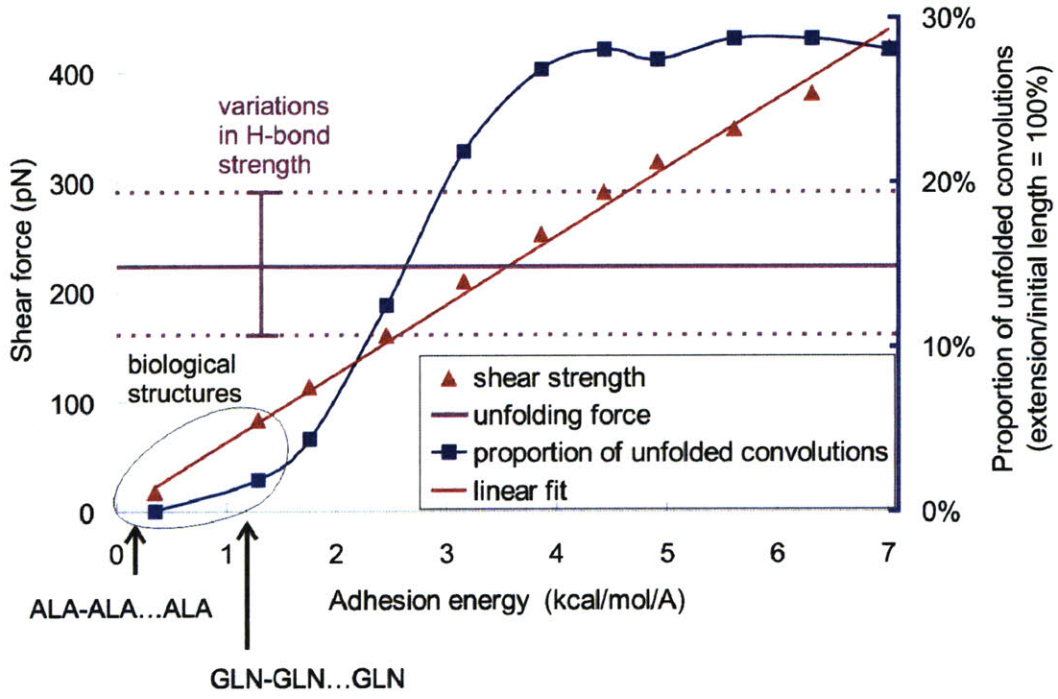

Figure 4.21 Influence of adhesion energy on the shear strength and the shear deformation mechanisms of two AHs in parallel. The figure depicts the curves of the shear strength and the proportion of unfolded convolutions over t he total number of convolutions (computed at a constant ratio of extension over initial length, 100%) versus adhesion energy per length. The fitted linear curve of the shear strength versus adhesion energy illustrates that the shear strength approximately equals the adhesion energy per length, which is in good agreement with our theoretical model (eq. (4.14)). The violet curve shows the constant value of the unfolding force of a single AH during a tensile loading experiment of same pulling velocity (about 200-220 pN [77, 81, 89], obtained from both atomistic, coarse-grained and theoretical studies, including an error bar to reflect variations of the H-bond strength). The relative magnitude of the shear strength compared to the unfolding force seems to explain the behavior of the curve of the proportion of unfolded convolutions over the total number of convolutions. For low adhesion energy per length values, the shear force is much lower than the unfolding force and thus we observe a very low proportion of unfolded convolutions. As the adhesion energy increases, the shear force becomes closer to the unfolding force and thus the probability of unfolding increases. Therefore we observe an increase of the proportion of unfolded convolutions. Once the shear force is above the unfolding force we reach a plateau because the two free segments (which are not anymore in parallel with another alpha-helix segment due to the shearing) are completely unfolded and are long enough to satisfy a 100% ratio extension over initial length so that the folded segments are not stretched any more.

Figure 4.21 presents the results of the influence of adhesion energy on the shear strength and the shear deformation mechanisms of two alpha-helices arranged in parallel (as shown in Figure 4.6 upper part). For our simulations we use 81 beads per alpha-helix strand and simulate the system at a temperature of 300 K. Figure 4.21 depicts the shear strength and the

proportion of unfolded convolutions versus adhesion energy per unit length. To quantify the number of unfolded alpha-helical convolutions, we compute the ratio of unfolded over total number of convolutions, evaluated at a constant ratios of extension over initial length (of 100%). The fitting curve of the shear strength illustrates that it approximately equals the adhesion energy per length (the fitting yields a factor of 0.9 between shear force and adhesion energy per length).

In a typical force strain curve of a shearing experiment of two alpha-helices, the shear strength is defined as the average force of the plateau regime. There are three distinct regimes. The first is an elasticity regime as the system begins to be sheared, followed by a plateau regime (which involves initially either molecular rupture or interprotein sliding, followed by interprotein sliding), and lastly a zero force regime once the two strands are completely separated from each other (where the force drops to zero rather quickly after the proteins have separated). These regimes can be seen in Figure 4.7, which will be discussed in depth below.

To provide an analysis of the shear strength of the protein assembly shown in Figure 4.7 upper part, we develop a simple theoretical model. Our model is based on the variation of the total "surface energy" of the system as the two alpha-helices are sheared. During the shearing experiment, once failure due to intermolecular shear initiates, the pulled alpha-helix slides relative to the fixed alpha-helix, and thus the surface of contact between the two strands decreases. Therefore the total surface energy of the system increases as the strands are being separated. Thus for a shear extension ΔL the work done by the pulling force f should equal the increase of "surface energy" of the system,

$$f \cdot \Delta L = \gamma \cdot \Delta L, \tag{4.13}$$

where γ is the adhesion energy per length as measured from full atomistic simulations (also referred to as "surface energy"). Eq. (4.13) leads to

$$f = \gamma . \tag{4.14}$$

Therefore the theoretical shear strength equals the adhesion energy per length, and our results are in good agreement with the theoretical model. Moreover, the theoretical value of the shear strength remains constant as we shear and thus agrees with the existence of a shear strength "plateau" regime as observed in the mesoscopic and atomistic simulations (Figure 4.7). This adhesion energy dependence of the shear strength, compared to the constant unfolding force

measured from tensile loading experiments of a single alpha-helix (about 200-220 pN [77, 81, 89]), explains the adhesion energy dependence of the proportion of unfolded convolutions over the total number of convolutions (computed at constant ratios of extension over initial length of 100%). This is discussed in more detailed below.

We now analyze the results obtained from our coarse-grained model (Figure 4.21). For low adhesion energy values (below 1.2 kcal/mol/Å), the shear force is much lower than the unfolding force, and thus the probability that unfolding of convolutions is observed is extremely low. In agreement with this notion, we observe a very low proportion of unfolded convolutions as the protein is shared (below 2%). As the adhesion energy increases, the shear force becomes closer to the critical unfolding force and thus the probability of unfolding increases. Therefore we observe an increase of the proportion of unfolded convolutions. The shear force exceeds the unfolding force for an adhesion energy of about 3.2 kcal/mol/Å (thus corresponding to 200-220 pN). Between 1.2 kcal/mol/Å and 3.2 kcal/mol/Å, we observe a strong increase of the proportion of unfolded convolutions. Above 3.2 kcal/mol/Å, the proportion of unfolded convolutions reaches quickly a plateau at a value of 28% whereas the shear force continues to increase linearly towards larger value. Once the shear force is above the unfolding force a plateau is reached because the two free segments, which are not anymore in parallel with another alpha-helix segment due to the shearing, are completely unfolded and are long enough to satisfy a 100% ratio extension over initial length so that the fold segments are not stretched anymore.

The behavior of the proportion of unfolded convolutions is also illustrated in Figure 4.22 in several snapshots of the deformation behavior. It presents the deformation of two AHs sheared at a ratio extension per initial length of 100%; from low to high adhesion energy structures: 1.29 kcal/mol/Å (panel A), 2.46 kcal/mol/Å (panel B), 4.43 kcal/mol/Å (panel C), 6.32 kcal/mol/Å (panel D). We note that the alpha-helix segments which are in parallel with another segment present no unfolding so that unfolded convolutions are only observed on the free alpha-helix segments.

To summarize, our results show that the predominant deformation mechanism depends strongly on the adhesion energy. As predicted from a simple theoretical model, the shearing force equals the adhesion energy per length. The magnitude of the shearing force compared to the unfolding force of a single alpha-helix (lower or higher) determine the predominant deformation mechanism of the shearing of two parallel alpha-helix. Thus, for low adhesion

energies, the predominant deformation mechanism is only sliding. As the adhesion energy increases the number of unfolded convolutions increases. For high adhesion energies (shear force above unfolding force), the predominant deformation mechanisms are both sliding and unfolding of convolutions.

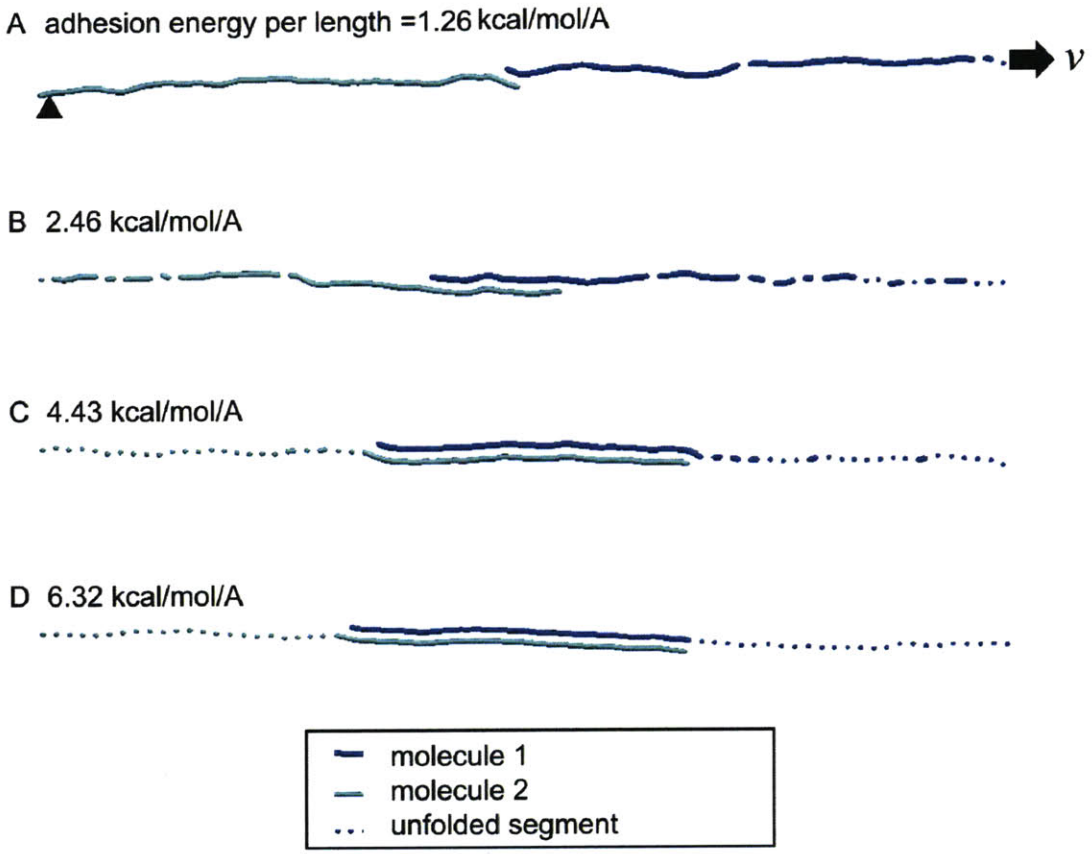

Figure 4.22. Deformation of two alpha-helices sheared at a ratio extension per initial length of 100%; from low to high adhesion energy structures: 1.29 kcal/mol/Å (panel A), 2.46 kcal/mol/Å (panel B), 4.43 kcal/mol/Å (panel C), 6.32 kcal/mol/Å (panel D). We use 81 beads per alpha-helix strand (432 Å) and a pulling velocity of 0.1 m/s. The results illustrate that the predominant deformation mechanism depends on the adhesion energy. For low adhesion energies, the predominant deformation mechanism is only sliding. The graphs clearly show that as the adhesion energy increases, the number of unfolded convolutions increases. For high adhesion energies, the predominant deformation mechanisms are both sliding and unfolding of convolutions. The alpha-helix segments which are in parallel with another segment present no unfolding so that unfolded convolutions are only observed on the free alpha-helix segments.

4.4.3 Discussion and conclusion

In this study we have applied a coarse-grained model of alpha-helical protein domains, which enabled us to simulate the dynamics of large systems over a large range of length- and time-scales and specific molecular properties of the constituting proteins. The main finding reported in this section is that interprotein sliding is a dominating mechanism that persists for a variety of geometries and realistic amino acid sequences (Figure 4.21). This suggests that alpha-helices remain largely intact during deformation of alpha-helix assemblies. Earlier experimental results on deformation of alpha-helical protein assemblies inside cells concur with this concept [94]. For very large adhesion values (in a range that does not correspond to adhesion properties associated with natural alpha-helices), our model predicts a change of deformation mechanisms where both sliding and unfolding occur concurrently. These strong adhesion values could be realized through the introduction of stronger interprotein bonding, e.g. via disulfide bonds that may provide a much greater shear resistance in excess of 1 nN. Future studies could explore this phenomenon.

The application of our coarse-grained model to structures of very different adhesion energies due to variations in amino acid sequences (see results shown in Figures 4.21-4.22) illustrates the unique ability of our mesoscale model to describe how adhesion properties and chemical structure influence strength properties and mechanisms of deformation of protein arrangements. Earlier attempts of describing the deformation mechanisms of protein domains with "coarse" mesoscale models have not yet included such level of detail, and have thus not been capable of describing how adhesion, chemical structure, size, geometry and deformation mechanisms and strength are linked.

4.5 Size effect of strength

4.5.1 For single AH arrangements

We illustrate an application of the mesoscale model to a study of the effect of the length of the alpha-helix protein domains on its strength properties, with protein lengths ranging from L=10.8 Å to L=4001.4 Å (that is, from 7 residues to 2668 residues), at a pulling rate of 0.1

m/sec. Figure 4.23 shows the strength obtained from our model for varying AH lengths. Here the strength of AHs is measured as the force value of the first rupture peak (see Figure 4.4). We observe two different regimes. For short AHs, the rupture strength decreases as the length increases. For long AHs, the rupture strength reaches an asymptotic value of approximately 180 pN (indicated with a dashed line in Figure 4.23). The transition between these two strength regimes occurs between 200 Å and 550 Å. We note that the simulation of the system with such long lengths would not have been possible with a full atomistic simulation (the full-atomistic simulation of a 70.2 Å protein structure took several weeks of computational time). For short AHs, the first force peak of the 2 convolution system (smallest one considered) is approximately 600 pN, and is thus almost four times as high than the 61 convolution structure (that is 329.4 Å, so around the transition of regimes), where the strength approaches 150 pN. We have used the short AH simulation results to fit an empirical equation of the form

$$f(L) = a\ln(L/L_0) + b, \tag{4.15}$$

where we find a = -117.97 pN, b = 864.6 pN and L_0 =1 Å.

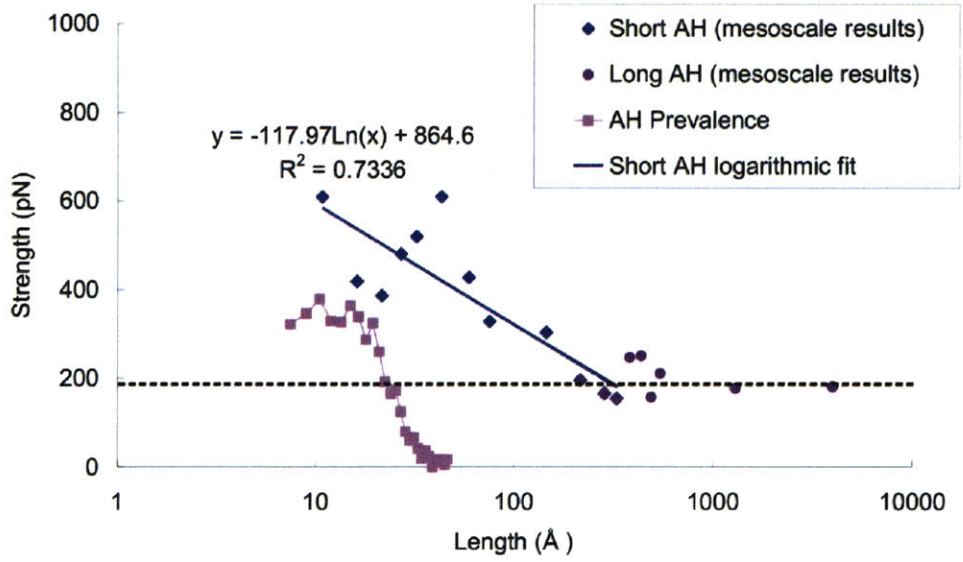

Figure 4.23. Size effects, strength, and prevalence of length of alpha-helices. This plot shows the strength properties of AHs with different lengths, ranging from L=10 Å to L=4000 Å (from 1 nm to 400 nm), at 0.1 m/sec pulling rate. The results illustrate two different regimes. For short AHs, the strength decreases as the length of the AH increases. The continuous line shows a logarithmic fit to the data obtained from mesoscale simulations (see eq. (4.15) for the equation and numerical fitting parameters). For longer AHs, the strength reaches an asymptotic regime of about 180 pN. The transition between these two regimes is between 200 Å and 550 Å. The plot of the prevalence over the alpha-helix length (distribution of AH single strand lengths found in biological proteins) illustrates that shorter alpha-helices are more prevalent [95]. This plot shows a correlation between the mesoscale results and the prevalence that is the most prevalent alpha-helices correspond to the strongest ones.

In summary, our model predicts that the strength of alpha-helices decreases as the length increases and then reaches an asymptotic regime for lengths higher than 200Å-550Å. The decrease behavior can be explained based on the fact that longer AHs contain more serial bonds, each of which can break with identical probability. Since failure of one convolution is sufficient to initiate failure of the entire system, we expect that longer molecules are weaker, as observed in our simulations and in [94]. This statement is also supported by the plot of the prevalence of length of alpha-helices that shows that the short alpha-helix are more prevalent in nature [89]. However, this analysis does not explain why we observe an asymptotic strength for long lengths and why the transition occurs between 200 Å and 550 Å. Perhaps, this asymptotic regime suggests that above a critical initial length of 200Å-550Å, an alpha-helical strand is stretched only on a segment of 200Å-550Å so that only the convolutions within this segment are applied a force and thus are likely to unfold. An other explanation

could be that the deformation of longer systems is smoother than shorter systems because they have much more interactions and thus the peak of failure is averaged. Thus the more bonds the closer we get to the average. Figure 4.24 seems to support this last hypothesis. It shows the force-strain curves of coarse-grained models of individual AH proteins for various initial length. We observe that the more bonds (or the higher initial length) the system has, the less significant the fluctuations are during unfolding. Moreover, the mean unfolding forces for all systems are very close to each other.

Lastly, we compare our simulation length effect results with the predictions of the theoretical model described in the first chapter of the Thesis. Although the logarithmic length dependence regime observed in the simulations agrees qualitatively with the theoretical model, it does not agree quantitatively. Indeed, eq.(3.14) gives a slope a = -34.5 pN which represents only 30% of the slope measured from the simulation fit. If we fit eq.(3.14) to the logarithmic dependence regime observed in simulation we find xb = 0.35 Å, that is also 30% of the parameter value used in the simulation. Thus this theoretical model on size-effect is not a good description of the system we study.

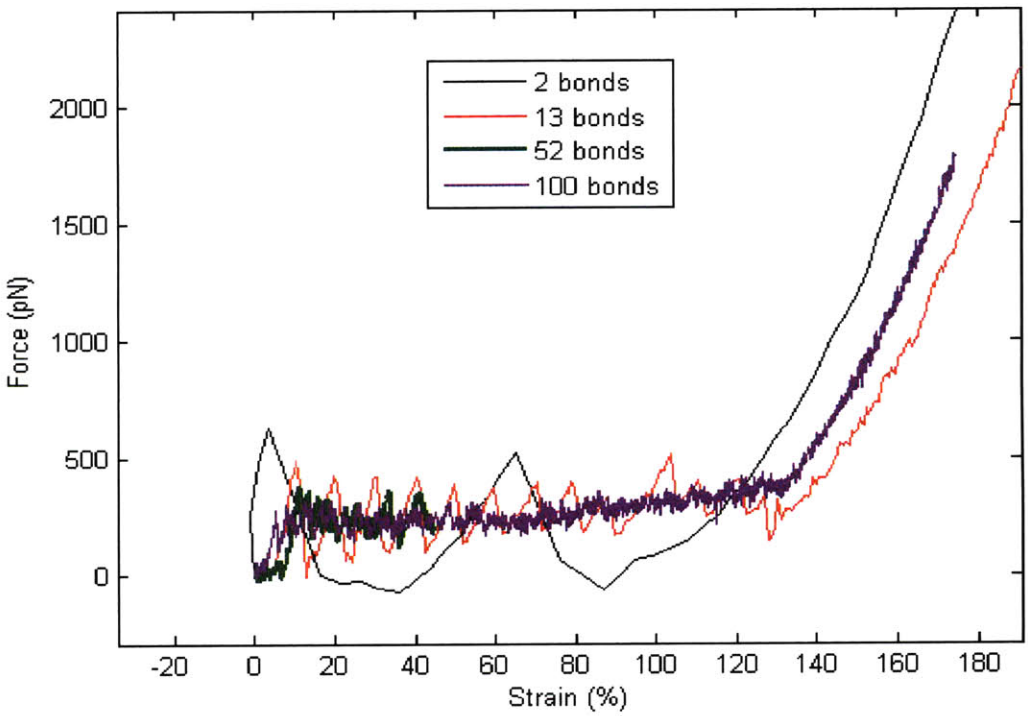

Figure 4.24. Force-strain curves of coarse-grained models of individual AH proteins for various initial length: 2 bonds, 13 bonds, 52 bonds and 100 bonds.

4.5.2 For parallel AH arrangements

We present another application of our mesoscopic model to study size effect of the shear strength of two parallel alpha helix proteins with the same three initial lengths: 21.6 Å, 70.2 Å and 432 Å. The schematic of the coarse-grained model for the shearing of two alpha-helix protein domains in parallel has been described earlier (Figure 4.7 upper part). Here we fit the adhesion energy per length γ to the value measured from full atomistic simulations for the reference case (GLN amino acid sequences: γ = 1.262 kcal/mol/Å, dimension of force). Figure 4.25(a) reminds us of the schematic of the coarse-grained model for the shearing of two alpha-helix protein domains in parallel.

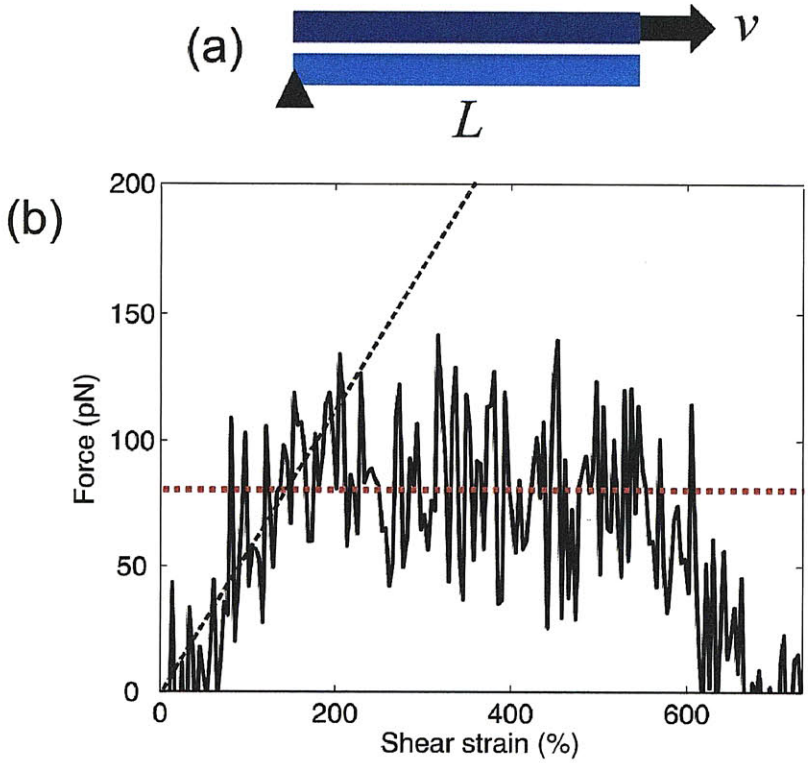

Figure 4.25. Shear strength of an assembly of two AHs. Panel (a) shows a schematic of the coarse-grained model setup for the shearing of two aligned AH protein domains. One strand has its end fixed and the second strand is pulled in the opposite direction at a constant velocity v =0.1 m/s. Panel (b) depicts the entire force strain curve for a shearing experiment on the mesoscopic model of two parallel alpha-helix protein domains (here we use 14 beads per AH strand, a pulling velocity of 0.1 m/s and a temperature of 300 K). The shear strain is defined as the ratio between the extension and the equilibrium distance between two AH strands. This plot shows three different regimes. First, an elasticity regime as we start shearing, then a plateau regime, which corresponds to the sliding of the pulled strand along the fixed strand and lastly a zero force regime once the two strands separate. The average of the force value over the plateau regime gives the mean shear strength (indicated with the horizontal red dashed line).

The extremity of one strand is fixed, and the second strand is pulled in the opposite direction at a constant velocity v =0.1 m/s. Figure 4.25(b) depicts the entire force strain curve for a shearing experiment on the mesoscopic model of two parallel alpha-helix protein domains (here we use a length of 14 beads for each AH strand, a pulling velocity of 0.1 m/s and a temperature of 300 K). The shear strain is defined as the ratio between the extension and the equilibrium distance between two AH strands. This plot shows three distinct regimes. First, an elasticity regime as the shearing begins, then a "plateau" regime which corresponds to the sliding of the pulled strand along the fixed strand, and lastly a zero force regime once the two strands separates. The average of the force value over the plateau regime provides an estimate for the mean shear strength.

Figure 4.26 shows snapshots during shearing of two AHs; for a short and a long AH structure. The predominant deformation mechanism is shearing; with only few unfolded convolutions appearing in the long case (red segments). This behavior is expected, since the shear resistance of 83 pN is lower than the rupture strength of AHs, which approaches values of more than 200 pN (see results in Figure 4.23).

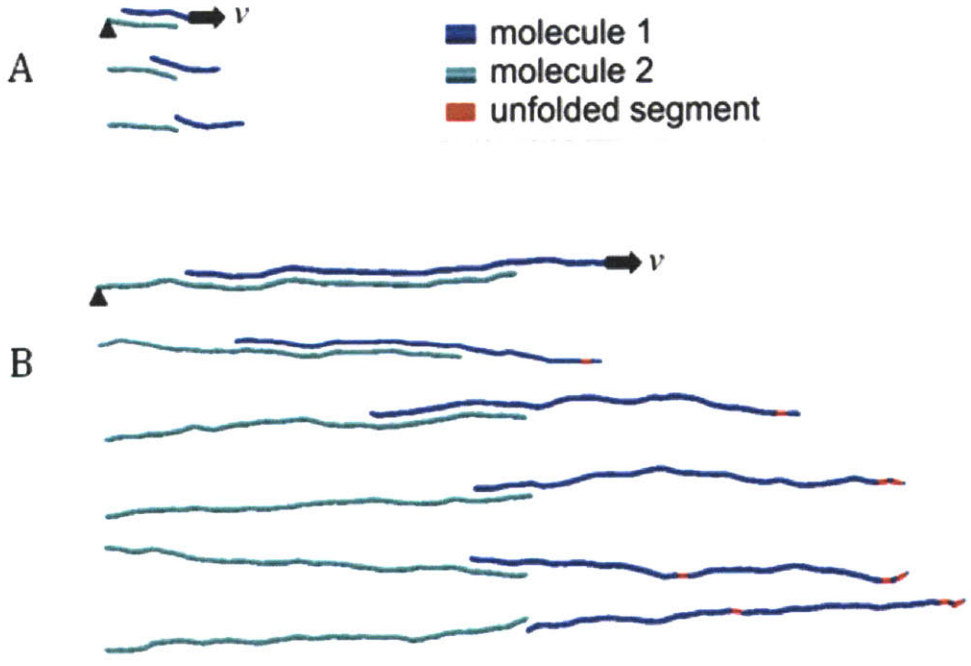

Figure 4.26. Dynamical mechanisms of shearing two AHs; for a short (panel A, length of 70.2 Å, 14 beads) and for a long (panel B, length of 432 Å, 81 beads) structure. The predominant deformation mechanism is sliding; with some unfolded convolutions appearing in the long case (visualized by red segments).

Figure 4.27 shows the shear strength properties of two parallel AHs obtained with three different lengths ranging from L=21.6 Å to L=432 Å, at 0.1 m/sec pulling rate. The shear strength is defined as the average force in the plateau regime observed in a typical force strain curve of a shearing experiment (see red straight line in Figure 4.25). The results illustrate the shear strength does not significantly depend on the length as opposed to the tensile strength. In average, the mean shear strength value is about 83 pN. The constant shear strength behavior may be explained by a simple theoretical model based on variation of surface energy (model described in eq. (4.15)).

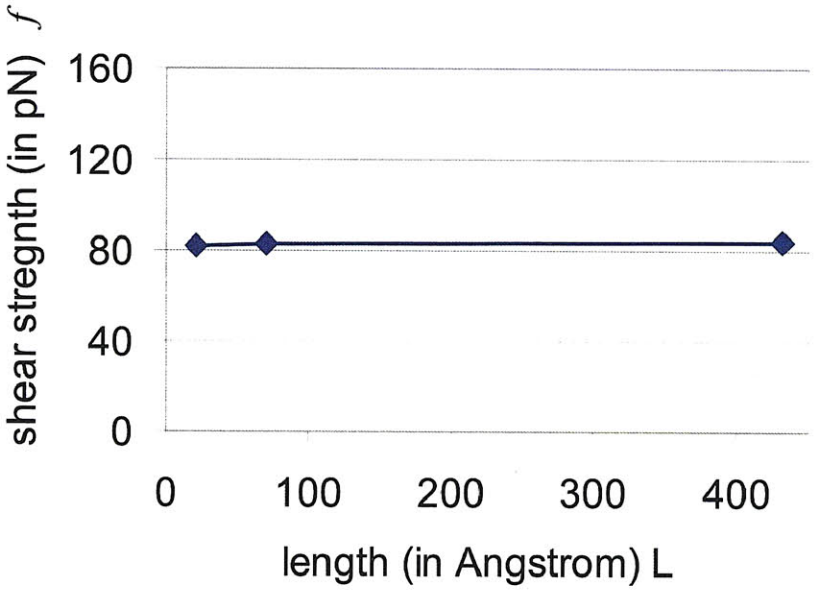

Figure 4.27. Shear strength properties of two parallel AHs with different lengths ranging from L=21.6 Å to L=432 Å, at 0.1 m/sec pulling rate. For different lengths, the plot shows the mean shear strength obtained from the force strain curve of the shearing experiment. The results illustrate the shear strength does not significantly depend on the length as opposed to the tensile strength, where the mean shear strength value is approximately 83 pN. The constant shear strength behavior may be explained by a simple theoretical model based on variation of surface energy (model described in eq. (4.14)). We note that for this range of lengths the shear strength is below the tensile strength (which is >200 pN, see Figure 4.23), and thus we do not observe unfolding of the whole AH strand as shear deformation occurs.

According to this model the theoretical shear strength equals 87.6 pN for γ = 1.262 kcal/mol/Å. The theoretical value of the shear strength approximately equals the shear strength value measured from simulations (83 pN) and indeed, as already discussed in detail in the previous section, this simple theoretical model seems to be relevant to describe the shear strength between parallel AHs. The small difference between the numerical values of theory and simulation may be explained by the fact that our simple theoretical model does not

include the effect of the pulling rate, which might shift the measured shear strength to larger values.

4.5.3 Discussion and conclusions

We have applied our mesoscale model to predict the strength of AH arrangements with different geometries and different lengths:

- Individual AH proteins: The study of length effect on strength shows two different regimes. For short AHs, we have shown a weakening effect as the molecule becomes longer. This could be explained by the larger number of serial coupled bonds, where failure of one bond is sufficient to initiate failure of the entire system (Figure 4.23). For long AHs, we have shown the existence of a transition from the strength decrease regime to a strength asymptotic regime. An explanation could be that the deformation of longer systems is smoother than shorter systems because they have much more interactions and thus the peak of failure is averaged. This might be one of the reasons why short single alpha-helices are more prevalent in nature than long ones, as they provide increased strength against failure. However our model cannot describe the mechanical role of coiled coil structures and protein folds into tertiary structures (AH protein domains). Such processes are known to provide much more stability and strength to the AH protein arrangements.

- Parallel AH proteins: The length does not affect significantly the shear strength and sliding is the prevalent deformation mechanism (versus unfolding).

The application to structures of different lengths and arrangements (see results shown in Figures 4.23-4.24 and Figures 4.25-4.27) illustrates the unique ability of our mesoscale model to describe how "structure" (that is, size, length, geometry) influences strength properties and mechanisms of deformation. Earlier attempts of describing the rupture mechanics of protein domains with "coarse" mesoscale models have not included such level of detail, and have thus not been capable of describing how structure and strength properties are linked. The insight into length-dependence of strength properties might be useful for the interpretation of experimental results and the comparison with theoretical and numerical models.

4.6 Discussion and conclusions

We have developed and calibrated (by fitting against full-atomistic MD results) a mesoscale model of AH protein arrangements, representing one convolution as a pair of mesoscale bead particles (Figure 4.2). This represents a reference system, which agrees well with the predictions from Bell's or Evans' models (Figures 4.4 and 4.5).

The coarse-grained model enables us to simulate the dynamics of large systems over a large range of length- and time-scales. The model is capable of reaching timescales of several microseconds and longer with a quantitative accuracy comparable with full atomistic MD simulations. Such relatively long simulations can be carried out within several days of computational time (on a single Intel Xeon CPU). In comparison, MD simulations of the dynamical behavior at fractions of microseconds can take weeks and months of computational time (even on a large parallelized simulation setup). Our coarse-grained model provides a considerable speedup while the model is still capable of describing the small- and large-deformation force-strain response characteristics as well as strength values quite accurately compared with atomistic simulations (see Figures 4.4-4.8).

We have carried out a systematic variation of the parameters in the mesoscale model and studied their impacts on the computational results and compared with commonly used protein strength models. We summarize the main findings below:

- Energy landscape parameters: the two most important contributions of this section are:
    - The reference system provided us with the starting point for a systematic variation of all relevant parameters, to identify how strength properties depend on the details of the energy landscape and under which conditions the Bell model fails. By systematically varying the energy landscape parameters, we have shown that the Bell model is a reasonable approximation to predict the mechanical strength properties as long as the model parameters are in a certain range: E_b between 9-18 kcal/mol (weak bond energy barrier) and x_r value above x_b value. (see, e.g. Figure 4.10).

Conversely, the Bell model breaks down when the model parameters are not in this range. We have also analyzed the dependence of the unfolding force on E_b and x_b (Figure 4.12) and confirmed the predictions from the Bell model.

- We have shown that the properties of the second, unfolded state of a AH convolution do not strongly influence the strength properties, provided that x_r , the distance between the unfolded state and the transition state, is higher than x_b , the distance between the folded state and the transition state (see, e.g. Figure 4.11). We have also shown that the energetic properties of the second unfolded state do not influence the rupture strength.

- Multi-timescale analysis: we have shown that under extremely small pulling velocities, the rupture strength of AH proteins approach an asymptotic value (Figure 4.15), where the strength does not depend on the pulling speed any more.

- Transducer stiffness: we have also reported an analysis of the effect of varying transducer stiffnesses on the strength properties, showing that stiffer transducers generally lead to increased strength properties (Figure 4.17).

- Bending stiffness: we have shown that under extremely small pulling speed, the model with change of bending stiffness is slightly stiffer than the model without this implementation (Figure 4.19).

- Bead mass: the results illustrate a very slight dependence of the rupture force on the bead mass (Figure 4.20).

Moreover, the coarse-grained model enables a computational engineering approach that allows us to rapidly screen a variety of conditions associated with alpha-helical protein properties, to identify critical conditions and mechanisms of how deformation is accommodated in this protein material. The application to structures of very different adhesion energies due to variations in amino acid sequences (see results shown in Figures 4.7 and 4.8) illustrates the unique ability of our mesoscale model to describe how adhesion properties and chemical structure influence strength properties and mechanisms of deformation of protein arrangements. Moreover, the application to structures of different lengths and arrangements (see results shown in Figures 4.23-4.24 and Figures 4.25-4.27)

illustrates the unique ability of our mesoscale model to describe how "structure" (that is, size, length, geometry) influences strength properties and mechanisms of deformation.

Earlier attempts of describing the deformation mechanisms of protein domains with "coarse" mesoscale models have not yet included such level of detail, and have thus not been capable of describing how adhesion, chemical structure, size, length, geometry and deformation mechanisms and strength are linked.

The mesoscale model has some limitations that could be addressed in future work. Compared to full atomistic simulations, the coarse-grained model is much less detailed and may only give a rough description of deformation mechanisms. Figures 4.7 and 4.8 illustrate that, in contrast to atomistic simulations, the mesoscale model do not show initial unfolding before sliding. Further, although we have included the effects of variations of the amino acid sequence on the intermolecular adhesion, we have not yet explored the effect of sequence variations on their rupture properties. Based on the analysis of the range of rupture strengths of alpha-helical proteins (see error bar associated with the variation of H-bond strength in Figure 4.21), our results should hold for a variety of cases as the shearing strength is lower than the lower bound of the error bar.

Potential applications of the coarse-grained alpha-helix protein domain model presented here could be further studies of length-scale effects on alpha-helix strength, elasticity and effects of hierarchical arrangements of alpha-helical based protein domains. Further studies could focus on larger variations of time-scales (e.g. to extend to cover very slow, experimental pulling speeds) and the development of similar formulations for coiled-coil proteins, larger scale protein folds with tertiary structures or larger-level protein materials.

Individual, isolated AHs are rarely found in biology. Therefore, further studies could be focused on applying the double-well potential to other structural proteins and filaments, or assemblies of AHs into larger-scale assemblies. The conclusions put forth here should be generically valid for a broader class of structural proteins. This is because the energy landscape structure as presented in Figures 4.1-4.2 is not unique to alpha-helices. Rather, other protein filaments featuring serially arranged domains that unfold under strain can be described based on the modeling framework proposed here (with appropriate parameterization of the model parameters listed in Table 4.1). In light of this, the choice of AHs in the studies reported can be considered as a model system that helped us to elucidate the generic behavior

reported can be considered as a model system that helped us to elucidate the generic behavior of a general class of protein filaments. Specifically, the insight into the length-dependence of strength properties (Figures 4.23-4.24), the effects of pulling speed (Figure 4.15) and transducer stiffness (Figure 4.17) might be useful for the interpretation of experimental results and the comparison with theoretical and numerical models for many other protein filaments.

5 Mechanics of cells and intermediate filament networks

Here we extend our modeling to describe larger-scale structures of protein materials, specifically focused on simulating the mechanical behavior of a cell. This study outlines the potential applications of a multiscale approach to describe more complex biological materials and systems.

The plan of this chapter is as follows. We begin with a presentation of the computational setup to study the mechanics of cell and intermediate filament networks. This includes a detailed description of the mesoscopic model formulation, the fitting procedure, as well as validation (Section 5.1). Section 5.2 is dedicated to the applications of the mesoscale cell model. It presents and discusses the computational results obtained to study the contribution of intermediate filaments to cell mechanical behavior.

5.1 2D mesoscale model formulation for cell

In the following sections we describe our multiscale simulation approach used to develop a mesoscale description of a cell with nucleus (*i.e.* a eukaryotic cell). The cell model is set up not based on a rigorous hierarchical multiscale approach; rather, it is based on a combination of experimental data and simulation data. The goal of developing this cell model is to study the mechanical contribution of cytoskeletal intermediate filaments (IFs) to cell structural integrity, stiffness and deformation in response to a given stress. Since we focus specifically on intermediate filament contributions (and how changes in their density influences the mechanical properties of cells), we do not include the other two cytoskeleton elements which are the microtubules (MTs) and the microfilaments (MFs), which also play an important role in cell structural integrity (see Figure 5.1A for a schematic of the structure of the cell's cytoskeleton). While this model is simplistic, it enables us to gain qualitative insight into the deformation behavior of cells and illustrates how the multiscale approach developed in this Thesis can be used to describe more complex biological structures. Our cell model is composed of the three main types of elements involved in cell structural integrity: Intermediate filaments, cell membrane and nucleus membrane.

Earlier modeling attempts at the cell level focused on cell mechanics and cell adhesion. A recent model discussed computer simulations of using a micropipette to attach and then detach a red blood cell on a flat substrate mediated by receptor-ligand binding [96]. Another continuum-mechanics-based hierarchical model has been proposed recently, applied to study the mechanics of mechanosensitive channel in bacteria Escherichia coli (E. coli) [97]. Other models used approaches such as tensegrity based simulations to describe cell mechanical properties [98]. However, thus far no model focused on the effects of intermediate filaments on cell mechanics, formulated based on discrete mesoscale elements, has been reported. This issue will be addressed in this Chapter, based on a simple model that enables us to simulate the effects of intermediate filament network density on cell mechanical properties.

Model formulation. Figure 5.1(B) shows the schematic of the cell model geometry (center and right region) inspired by experimental pictures (left region, reprinted from [99]). The experimental view depicts the geometry of Madin-Darby canine kidney (MDCK) epithelial cells, including the IF network within the cells (scale bar is approximately 25 μm). IFs extend from the nucleus to the cell membrane. The schematic of the cell model geometry represents the three different coarse-grained models used respectively for intermediate filaments, cell membrane and nucleus membrane. Each model try to capture the main features of the cell element that they represent, as far as geometry and mechanical properties are concerned. The cell model diameter is 10 μm. Figure 5.1(C) presents the schematic of the loading geometry of a single cell, implemented by fixing the left end of the cell (black circle) and pulling on the right end of the cell (red circle). This setup mimics an optical tweezers experiment of cell stretching.

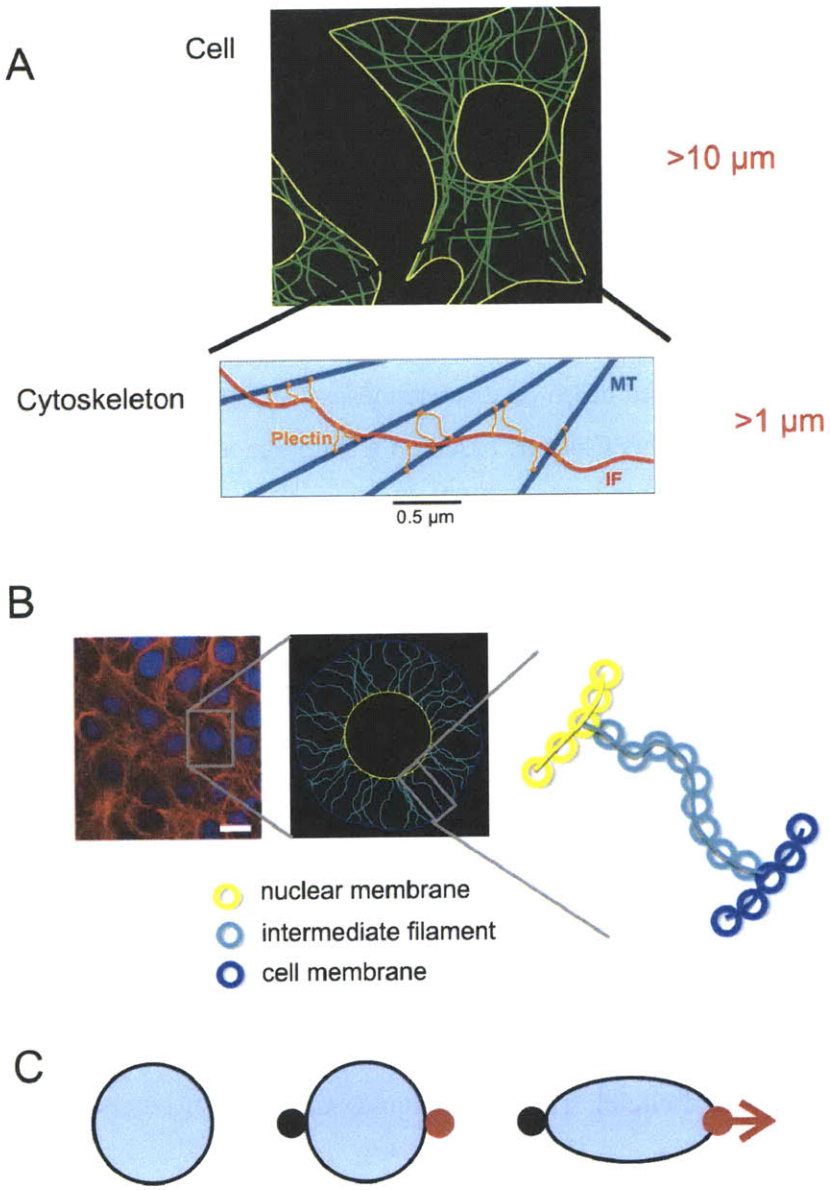

Figure 5.1. Simple schematic of the cell and the underlying cytoskeletal structure and loading geometry used for the single cell stretching experiments. Panel A: Structure of the cell's cytoskeleton. The cell's cytoskeleton contains actin filaments (=microfilaments, mostly found beneath the cell membrane), microtubules, and intermediate filaments. Here we only show intermediate filaments and microtubules in the blow-up and their cross-linking structure via plectin proteins (plectin proteins cross-link all three major cytosketal constituents). Panel B: Schematic of the cell model geometry (right region) inspired by experimental pictures (left region). The experimental view depicts the geometry of Madin-Darby canine kidney (MDCK) epithelial cells, including the IF network within the cells (scale bar is approximately 25 μm). IFs extend from the nucleus to the cell membrane. The cells were grown on collagen-coated silastic membranes. Cells were fixed and stained for immunofluorescence (red=keratin IFs, blue=DNA). The schematic of the cell model geometry represents the three different coarse-grained models used respectively for intermediate filaments, cell membrane and nucleus membrane. Each model try to capture the main features of the cell element that they represent, as far as geometry and mechanical properties are concerned. The cell model diameter is 10 μm. The experimental picture is reprinted from [99]. Panel C: Schematic of the loading geometry of a single cell, implemented by fixing the left end of the cell (black circle) and pulling on the right end of the cell (red circle). This setup mimics an optical tweezers experiment of cell stretching.

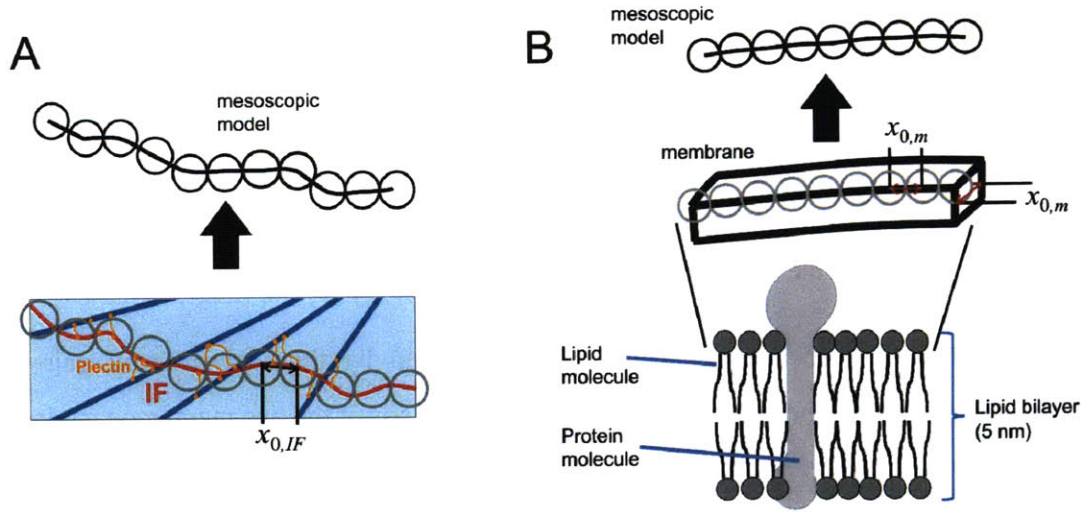

Figure 5.2. Schematic of the coarse-graining procedure, representing intermediate filaments and membranes by a mesoscale bead-spring model. The values of the model parameters (see parameter identification Sections 2.3.1-2.3.3 and Tables 1-3) are specific to each cell element, and derived from geometrical analyses, experimental and atomistic simulations. An intermediate filament is a hierarchical arrangement of alpha-helical proteins that has a 1D shape and thus can be easily described by a bead chain with $x_{0,IF}$ as the intermediate filament equilibrium bead distance. Membranes are lipid bilayers with embedded membrane-associated proteins. Because of its 2D shape, representing a membrane by a 1D bead chain is more approximate. For membranes, the equilibrium bead distance $x_{0,m}$ determines also the width of the membrane that is represented by one bead.

Figure 5.2 depicts a schematic of the geometry and the coarse-graining procedure, representing intermediate filaments and membranes by mesoscale bead models. The values of the model parameters are specific to each cell elements and derived from geometrical analyses, experimental and atomistic simulations (the fitting is described later in the text). An intermediate filament is a highly hierarchical arrangement of alpha-helical proteins that has a 1D shape and thus can be well described by a bead chain with $x_{0,\mathrm{if}}$ as the IF equilibrium bead distance. Membranes are lipid bilayers with membrane-associated proteins. Because of its 2D shape, representing a membrane by a 1D bead chain is more approximate. For membranes, the equilibrium bead distance $x_{0,\mathrm{m}}$ determines also the width of the membrane represented by one bead.

The mesoscale approach is similar to the one used for the mesoscale model of single AHs. The beads in the cell mesoscale model interact according to an intermolecular multibody potential, developed to reflect the key physical properties of intermediate filaments and membranes including stretching and bending.

The mathematical expression for the total energy of the system is given by

$$E\left(\vec{X}\right) = E_T + E_B + E_I, \tag{5.1}$$

where $\vec{X}$ denotes the positions of all beads, E_T the total tensile energy and E_B the total bending energy and E_I the total intermolecular interaction energy.

The total intermolecular interaction energy E_I is given by the sum over all pair-wise interactions between the beads of the IF models and the beads of the two membrane models,

$$E_I = \sum_{pairs} \phi_I(x) \ . \tag{5.2}$$

The role of the intermolecular potential ϕ_I is to avoid IF bead penetration through the membranes. Thus ϕ_I is described by a repulsive Lennard-Jones potential,

$$\phi_I(x) = 4 \cdot \varepsilon \cdot \left[\left(\frac{\sigma}{x}\right)^{12} - \left(\frac{\sigma}{x}\right)^{6} \right] \qquad x < x_c, \tag{5.3}$$

with ε the energy minimum of the Lennard-Jones potential, x the distance between mesoscale particles, x_c the cutoff distance and σ the zero-crossing distance.

The total tensile energy E_T is given by the sum over all pair-wise interactions and the total bending energy E_B by the sum over all three-body interactions, where

$$E_T = \sum_{pairs} \phi_T(x) \text{ and } E_B = \sum_{triplets} \phi_B(\theta) \tag{5.4}$$

We define specific interparticle potential energy expressions for each of the three cell elements.

Intermediate filament. We approximate the nonlinear force-extension behavior of IFs under tension by a multilinear model. This multilinear model is a combination of six spring constants $K_T^{(i)}$ (i = 1..6), which are turned on at specific values of molecular stretch. A similar model has been used successfully in earlier studies of fracture in crystalline model materials [100] and provides an effective way to describe the nonlinear constitutive behavior based on computationally effective, simple piecewise harmonic potential functions. Based on this model, the tensile force between two bead particles is described as:

$$F_T(x) = -\partial\phi_T(x)/\partial x, \tag{5.5}$$

(the energy function ϕ_T is given by integrating the force $F_T(x)$ over the radial distance), where

$$\frac{\partial\phi_T}{\partial x}(x) = H(x_{break} - x)\begin{cases} K_T^{(1)}(x - x_0) & x > x_1 \\ R_1 + K_T^{(2)}(x - x_1) & x_1 \le x < x_2 \\ R_2 + R_1 + K_T^{(3)}(x - x_2) & x_2 \le x < x_3 \\ R_3 + R_2 + R_1 + K_T^{(4)}(x - x_3) & x_3 \le x < x_4 \\ R_4 + R_3 + R_2 + R_1 + K_T^{(5)}(x - x_4) & x_4 \le x < x_5 \\ R_5 + R_4 + R_3 + R_2 + R_1 + K_T^{(6)}(x - x_5) & x_5 \le x \end{cases} . \tag{5.6}$$

In eq. (5.4), $H(x_{break} - x)$ is the Heaviside function $H(a)$, which is defined to be zero for $a < 0$, and one for $a \ge 0$. The parameters $R_1 = K_T^{(1)}(x_1 - x_0)$, $R_2 = K_T^{(2)}(x_2 - x_1), R_3 = K_T^{(3)}(x_3 - x_2)$, $R_4 = K_T^{(3)}(x_3 - x_2)$ and $R_5 = K_T^{(3)}(x_3 - x_2)$ are calculated from force continuity conditions.

The bending energy of a triplet of three bead particles is given by

$$\phi_B(\theta) = \frac{1}{2}K_B(\theta - \theta_0)^2 \tag{5.7}$$

with K_B the bending stiffness parameter relating to the IF bending stiffness EI through $K_B = 3\frac{EI}{x_0}$. Note that the IF bending stiffness EI is related to the IF persistence length L_p through $L_p = \frac{EI}{k_b T}$ where k_b is the Boltzmann constant and T the temperature.

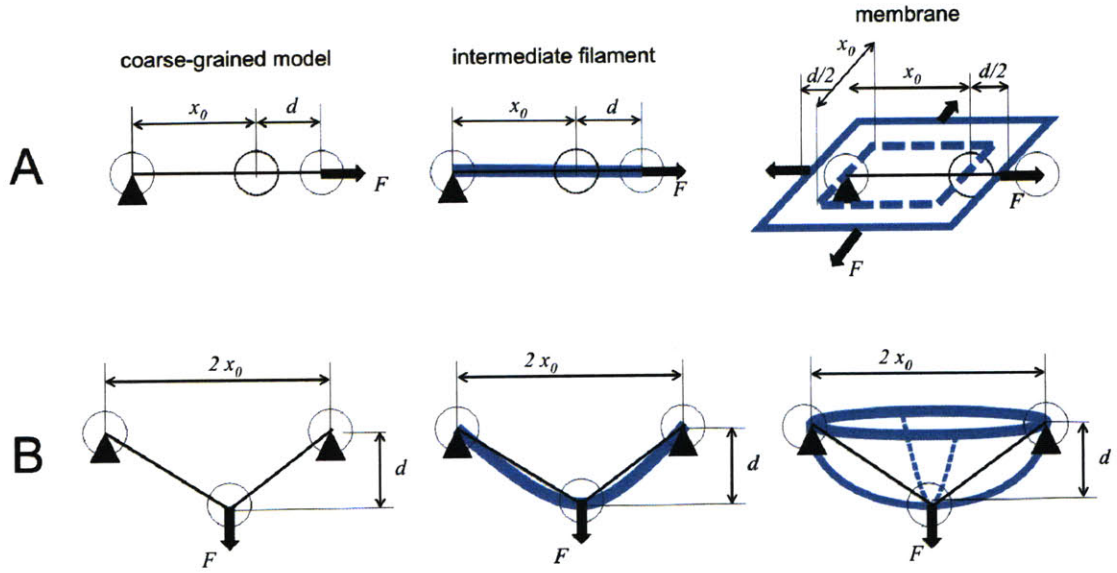

Figure 5.3. Mechanical loading experiments used to fit tensile (panel A) and bending (panel B) potential parameters for IF and membrane coarse-grained models. For IFs, tensile parameters are fitted so that the force strain curve of a stretching mesoscale simulation is identical to the one obtained from a stretching experiment. The bending parameters are obtained by identifying the bending energy of a three bead mesoscale model to the bending energy of IF of same length during a three point bending experiment for small deformations d. For membranes, the parameters of the tensile pair-wise interparticle potential with an equilibrium distance x_0 are obtained by identifying a tensile harmonic potential of a two bead model with an extension d to the tensile energy of a membrane with an area x_0 × x_0 during an experiment with uniform extension equal to d. The parameters of the bending three-body potential are obtained by identifying the bending energy of a three bead mesoscale model to the bending energy of a round membrane with radius x_0 during a bending experiment for small deformations d.

Cell and nucleus membranes. As mentioned earlier, modeling a 2D membrane as a 1D bead chain is rather approximate. For instance, membranes can be stretched along multiple directions at the same time (for instance along x and y axis if the membrane is in the x-y plane) whereas a 1D bead chain cannot. Thus we need to make assumptions on how to define tensile and bending potentials for the bead model in order to link potential parameters to the key mechanical properties of the membranes determined by experiments or simulations. Figure 5.3 shows the real loading experiments used to fit tensile and bending potential parameters for IF and membrane coarse-grained models.

The tensile energy of a pair-wise interaction is given by

$$\phi_T(x) = \frac{1}{2}K_T(x - x_0)^2 \tag{5.8}$$

with the tensile stiffness parameter K_T relating to the area expansion modulus K_e of the membrane through $K_T = 4K_e$. This relation is derived by identifying the tensile harmonic

potential of a two bead model with an extension d to the tensile energy of a membrane with an area $x0 \times x0$ during an experiment with a uniform extension equal to d. Below are the details of this derivation.

The tensile energy of the membrane is given by

$$\phi_T(\Delta A) = \frac{1}{2} K_e \left(\frac{\Delta A}{A_0} \right)^2 A_0, \tag{5.9}$$

where ΔA is the expansion of the initial membrane area $A_0 = x_0^2$.

For uniform biaxial small strain $\varepsilon = \varepsilon_1 = \varepsilon_2 = \frac{d}{x_0}$, we have

$$\phi_T(\Delta A) = \frac{1}{2} K_e (2\varepsilon)^2 A_0 = 2K_e d^2 \tag{5.10}$$

thus, by identifying eq. (5.8) to eq. (5.10) given that $x - x_0 = d$, we get

$$K_T = 4K_e \ . \tag{5.11}$$

The bending energy of a triplet of three bead particles is given by

$$\phi_B(\theta) = \frac{1}{2} K_B (\theta - \theta_0)^2 \tag{5.10}$$

with K_B relating to the bending stiffness of the membrane K_b through $K_B = \pi K_b$.

This relation is derived by identifying the bending energy of a three bead mesoscale model to the bending energy of a round membrane with radius x_0 during a bending experiment for small deformations d. Below are the details of this derivation.

The bending energy of the membrane with an initial area $A_0 = \pi x_0^2$ is given by

$$\phi_B(\theta) = \frac{1}{2} K_b \left(\frac{2}{R} \right)^2 A_0 \tag{5.13}$$

where $\frac{1}{R}$ is the curvature caused by the deflection d (the two principal curvatures are identical).

For small deflections, $\frac{1}{R} = \frac{d}{x_0^2}$ and $\theta - \theta_0 = 2\frac{d}{x_0}$ thus by identifying eq. (5.12) to eq. (5.13) we obtain

$$K_B = \pi K_b \,. \tag{5.14}$$

Model parameter identification. All parameters in the cell mesoscale bead model except the nucleus bending stiffness are determined from geometrical analysis, experimental and full atomistic simulation results. In the next few paragraphs we describe the parameter fitting approach for each of the three cell model elements and summarize the entire set of parameters in three tables.

Intermediate filaments. We choose x_0 = 0.2492 um per bead (equilibrium bead distance), providing significant computational speedup while maintaining a sufficiently fine discretization of the intermediate filament (equilibrium bead distance is taken lower than the IF persistence length which is about 1 μm [101]. It leads to a bead particle mass m= 8723077 amu. All parameters of the tensile interaction in eq. (5.6) are fitted to reproduce the force strain curve behavior obtained using both experimental results [102] and full atomistic results [101]. Figure 5.4 depicts the force-strain curve for alpha-helices as reproduced by the mesoscale bead model.

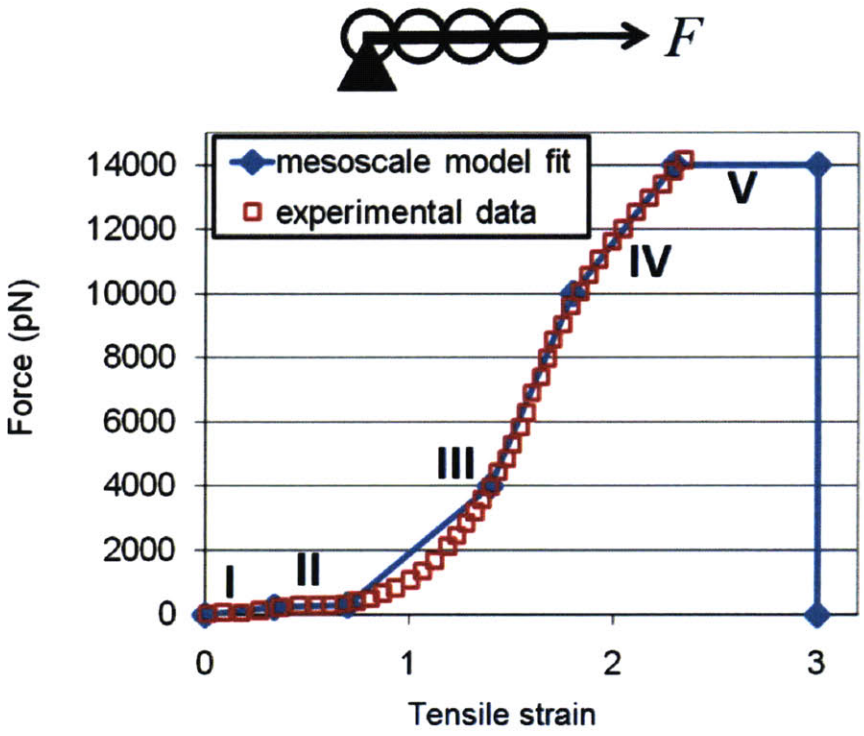

Figure 5.4. Characteristic force-strain curve for stretching an intermediate filament. The square plot depicts the experimental results for a single intermediate filament (from hagfish slime [103] assuming a 10nm IF diameter). The solid line represents the force-strain curve for single intermediate filament as reproduced by the IF mesoscale model. It is derived from the experimental results and very recent full atomistic simulations on IF tetramers [104]. The curve shows five major deformation regimes (regions I, II, III, IV and V).

The curve shows five major deformation regimes. The first is the low stiffness regime (region I with stiffness $K_T^{(1)}$) which is followed at a strain of 34% (this determines the onset point for the second regime, x_1) by an even lower stiffness plateau (region II with stiffness $K_T^{(2)}$). Region III with a stiffness $K_T^{(3)}$ and then $K_T^{(4)}$ (transition strain of 140%, which determines the transition point within the third regime, x_3) corresponds to a high rising stiffness regime, starting at a strain of 70% (this determines the onset point of the second regime, x_2). Experimental analysis [103] and very recent full atomistic results [103] suggest that regions II and III correspond to the disruption of coiled-coil alpha helices in IFs, and the formation of stable beta-sheet crystallite. Starting from the strain of 180% (this determines the onset point of the fourth regime, x_4), Region IV corresponds to a dropping stiffness regime leading to a zero stiffness plateau regime at a strain of 230% (this determines the onset point of the fifth regime, x_5) and a force of 14 nN (region V). The existence of this last stiffness 'yield' regime is suggested by full atomistic simulations carried out on single IF tetramer [103]. The breaking strain is 300%. The bending stiffness parameter K_B is obtained from experimental

results on IF persistence length. IF persistence length is about 1 μm [2] thus we determine $K_B = 3.47\,\text{kcal/mol/rad}^2$. The value of the equilibrium angle θ_0 is 180 degrees, based on the geometry of the IF structure. The entire set of parameters for the IF mesoscale model is summarized in Table 5.1.

Parameter and units	**Numerical value**
Equilibrium bead distance x_0 (in Å)	2492
Critical distances x_1, x_2, x_3 x_4 and x_5 (in Å)	3339, 4236, 5981, 6978, 8224
Tensile stiffness parameters $K_T^{(1)}$, $K_T^{(2)}$, $K_T^{(3)}$, $K_T^{(4)}$, $K_T^{(5)}$ and $K_T^{(6)}$ (all in kcal/mol/Å^2)	0.001955, 0.0005621, 0.01528, 0.04336, 0.02313, 0
Bond breaking distance x_{break} (in Å)	9968
Equilibrium angle θ_0 (in degrees)	180.00
Bending stiffness parameter K_B (in kcal/mol/rad^2)	3.47
Mass of each mesoscale particle m (in amu)	8723077

Table 5.1. Summary of parameters of the IF mesoscale model, derived from geometrical analysis, experimental and full atomistic simulation results.

Parameter and units	**Numerical value**
Equilibrium bead distance x_0 (in Å)	1571
Tensile stiffness parameter K_T (in kcal/mol/ Å ^{2})	1.44
Equilibrium angle θ_0 (in degrees)	180.00
Bending stiffness parameter K_B (in kcal/mol/rad^2)	144
Mass of each mesoscale particle m (in amu)	171880665

Table 5.2. Summary of parameters of the mesoscale model for the cell membrane, derived from geometrical analysis, experimental and full atomistic simulation results.

Cell membrane. We choose $x_0 = 0.1571$ um per bead (equilibrium bead distance), providing significant computational speedup while maintaining a sufficiently fine discretization of the

membrane. This value combined with data on the membrane mass composition [105] lead to a bead particle mass m = 171880665 amu. We note that the thickness of a cell or nucleus membrane is about 5 nm. As described in eq. (5.11), the tensile stiffness parameter K_T is linked to the area expansion modulus K_e of the membrane , which can be obtained from experimental measurements. Typical K_e values lie in the range of 0.1-1 N/m for various types of lipid bilayers. For instance K_e is about 0.45 N/m for red blood cell (RBC) membranes [106]. Thus we identify K_T as being about 1 N/m, that gives us $K_T = 1.44 \text{kcal/mol/A}^2$. As described by eq.(5.14), the bending stiffness parameter K_B is linked to the bending stiffness K_b of the membrane, which can be obtained from experimental measurements. Typical K_b values lie in the range of 10^{-19} Nm for RBCs or lipid bilayers [107]. This value is larger, on the order of $1\text{-}2\text{x}10^{-18}$ Nm [2] for other cell types (e.g., neutrophils, endothelial cells) that possess a more extensive cytoskeleton network. Thus we identify K_B as being about $2\text{x}10^{-18}$ Nm, that gives us $K_B = 144 \text{kcal/mol/rad}^2$. The value of the equilibrium angle θ_0 is 180 degrees, which corresponds to a relaxed membrane structure with no curvatures. The entire set of parameters for the mesoscale model of the cell membrane is summarized in Table 5.2.

Parameter and units	**Numerical value**
Equilibrium bead distance x_0 (in Å)	1571
Tensile stiffness parameter K_T (in kcal/mol/ Å ^{2})	1.44
Equilibrium angle θ_0 (in degrees)	180.00
Bending stiffness parameter K_B (in kcal/mol/rad^2)	144000
Mass of each mesoscale particle m (in amu)	171880665

Table 5.3. Summary of parameters of the mesoscale model for the nucleus membrane, derived from geometrical analysis, experimental and full atomistic simulation results.

Nuclear membrane. We choose x_0 = 0.1806 um per bead. This value combined with data on the membrane mass composition [99] lead to a bead particle mass m = 227312178 amu. Similar to the cell membrane, we identify the tensile stiffness parameter K_T as being about 1 N/m, that gives us $K_T = 1.44 \text{kcal/mol/A}^2$. However, we use a K_B value one thousand times higher than the one used for cell membrane, that means $K_B = 144000$ kcal/mol/rad^2. We use a much stiffer value to avoid significant deformations of the nucleus while stretching the cell as

we observe in our simulations. This correction may illustrates the intrinsic errors made when we represent a 3D nucleus structure as a simple 1D bead chain. The value of the equilibrium angle θ_0 is 180 degrees. The entire set of parameters for the mesoscale model of the nucleus membrane is summarized in Table 5.3.

Now we can calibrate the Lennard-Jones potential parameters introduced in eq. (5.3) to avoid IF bead penetration through the membranes. We choose a very high energy minimum value ε = 10000 kcal/mol/Å^2 and a short cutoff distance x_c = 1350 Å to have a strong repulsion localized near the membranes.

Parameter and units	**Numerical value**
Energy minimum ε (in kcal/mol/ Å^2)	10000
Zero-crossing distance σ (in Å)	1200
Cutoff distance x_c (in Å)	1350

Table 5.4. Summary of parameters of the repulsive Lennard-Jones potential between the IF beads and the membrane beads to avoid IF bead penetration through the membranes.

Lastly, since each bead of the cell mesoscale model represents a large group of atoms, their significant masses enable us to choose a relatively large timestep for fast computing. Thus we choose a timestep of 10000 fs, which is still below the characteristic time constant $\tau = \sqrt{\frac{k}{m}}$.

Model geometry and boundary conditions. The cell model geometry is an approximation of the geometry of an eukaryotic cell, in particular cells such as epithelial cell or fibroblast that are rich in IFs (keratin IFs and vimentin IFs, respectively). Figure 5.5(A) shows the geometry of Madin-Darby canine kidney (MDCK) epithelial cells, including the IF network within the cells. IFs extend from the nucleus to the membrane. Our cell coarse-grained model represents the main geometrical features. We choose to represent a cell of small size (such as fibroblast) in order to allow fast computing. Thus we build a round cell coarse-grained membrane with a diameter of 10 μm and a round nucleus coarse-grained membrane with a diameter of 4.6 μm so that in reality the volume of the nucleus is about 10% of the total cell volume. We peripherically distribute radial coarse-grained IFs that are attached from one extremity to the nucleus and from the other one to the membrane. The tensile bonds between IF extremities and membranes are chosen the same as intra IF tensile bonds. There are no angle bonds define

for these junctions. Each IF is made of 12 beads. All IFs are given a slightly slack behavior after equilibration by setting the initial inter bead distances to a smaller value than the bead equilibrium distance. Figure 5.5(C-D) depicts a schematic of the coarse-grained cell geometry under static (after equilibration) and strained conditions. This system has 40 IFs and one thousand particles in total. All simulations are carried out at 300 K in a NVT ensemble (constant temperature, constant volume, and constant number of particles). We constrain the simulations in two dimension (x and y axis). All simulations are 2D simulations. We equilibrate the structure during 4 000 ns at a temperature of 300K by constraining lateral motions (along y axis) of the top and bottom extremities. Then we pulled along the y axis on each of these two extremities at a pulling velocity of 0.1 m/s and measure the force-strain curve. The simulations typically run for less than an hour.

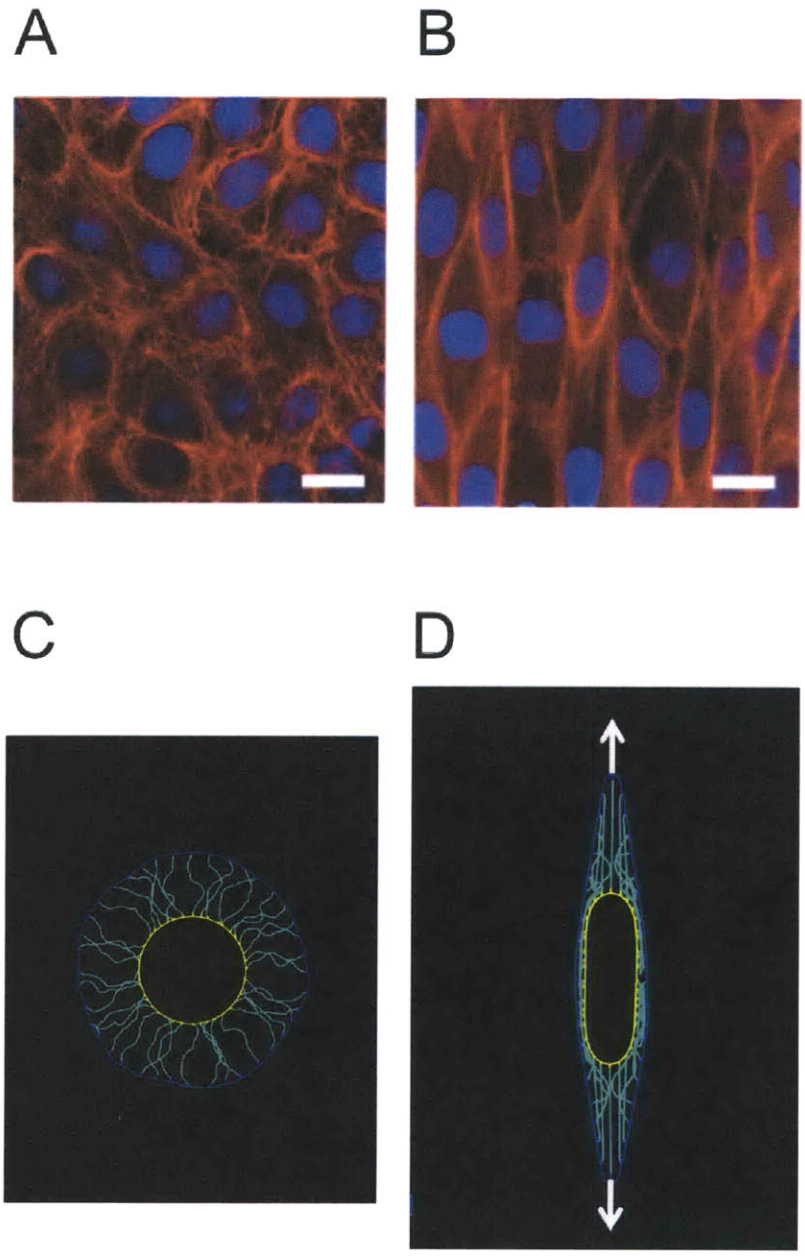

Figure 5.5. Cell model geometry, boundary conditions and illustration of the role of intermediate filament network for maintaining cell integrity at large deformation. Panels A and B: Intermediate filament network in MDCK cells (scale bar is approximately 25 μm) under static and strained conditions. Panels B and C: Visualization of the coarse-grained cell geometry under static and strained conditions. Upper images in panels A and B reprinted from [99].

5.2 Contribution of intermediate filaments to cell mechanical behavior

Figure 5.5(B) depicts a view of MDCK cells stretched uniaxially at about 75%. This snapshot illustrates the role of IFs for maintaining cell integrity at large deformations. Here we use our cell model to study in depth the roles of IF network within cells. Figure 5.5(D) shows a snapshot of the cell model stretched at about 50%.

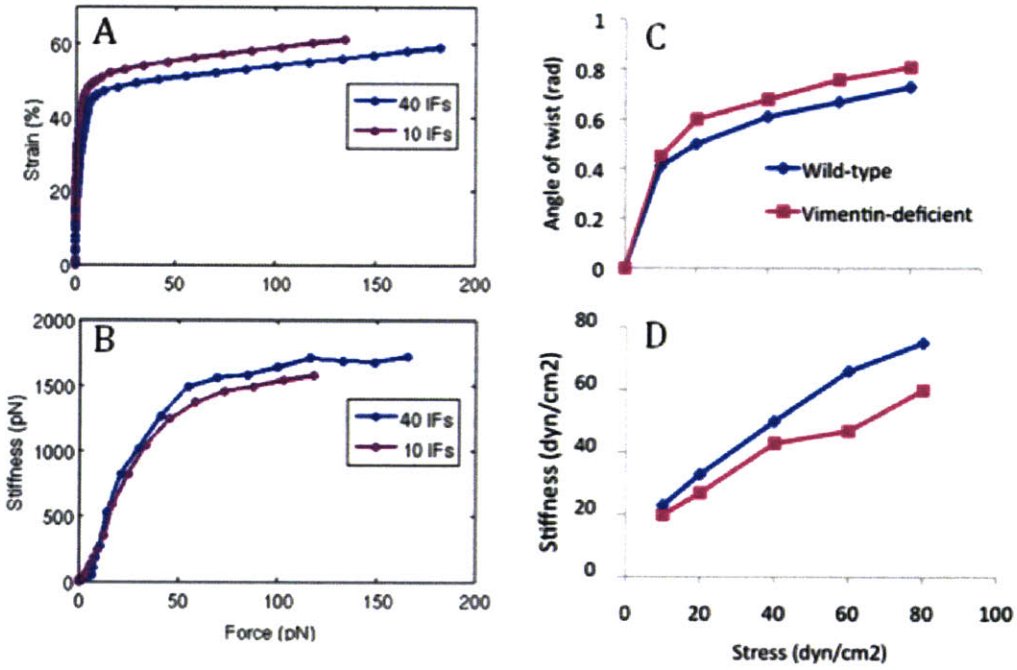

Figure 5.6. Mechanical response of the cell under tension for several intermediate filament (IF) densities, as predicted by the mesoscale model (Subplots A and B) and in comparison with experimental results (Subplots C and D, adapted from [89]). The experimental measurements were done using magnetic cell twisting cytometry. Stresses ranging from 10 to 80 dyn/cm^2 were applied through Arg-Gly-Asp coated ferromagnetic beads. Cells were plated on collagen I-coated dishes for 4h before mechanical measurements. Subplots A and B show respectively strain-force curves and stiffness-force curves for cell models with high density (blue curve, with 40 IFs) and low density (pink curve, with 10 IFs) of IFs. Subplots C and D show respectively angle of twist (angular strain)-stress curves and stiffness-stress curves for adherent wild-type and intermediate filament-deficient fibroblasts. Note that as IF density increases, the cell model exhibits greater stiffening and greater stiffness. These results are consistent with the experimental measurements and agree with the results from our mesoscale cell model.

First we present the results of the mechanical response of the cell under tension for several intermediate filament (IF) densities, as predicted by the mesoscale model. The results are shown in Figure 5.6 (subplots A and B) and in comparison with experimental results (subplots C and D [108]) . The experimental measurements were done using magnetic cell twisting cytometry. Subplots A and B show respectively strain-force curves and stiffness-force curves

for cell models with high density (40 IFs) and low density (10 IFs) of IFs. Subplots C and D show respectively angle of twist (angular strain)-stress curves and stiffness-stress curves for adherent wild-type and intermediate filament-deficient fibroblasts. The two methods plot different quantities that are qualitatively comparable. Indeed, an angle of twist is equivalent to a deformation and thus to a strain. Moreover, a stress is equivalent to a force. Thus both methods plot a "deformation-force" curve and a "stiffness-force" curve. We observe that as IF density increases, the cell model exhibits greater stiffening and greater stiffness. These results are consistent with the experimental measurements and agree qualitatively with our results.

As the next step we analyze the dynamics of cell deformation during a stretching experiment. Figure 5.7 depicts the snapshots of the cell model with 40 IFs at different applied tensile strains. Region I shows the initial model geometry. The system has 40 radial intermediate filaments uniformly peripherically distributed. Region II depicts the geometrical configuration of the model after equilibration (4000 ns). Note that IFs are slack and membranes have small fluctuations. Regions III, IV and V show snapshots of the cell model at various stretching strains (respectively 24%, 48% and 59%). As we pull the cell, the intermediate filament network is stretched in the areas between the pulled extremities and the nucleus, and compressed (bended) in the cell lateral areas. At small strains (region III), the IF network is deformed as described above and the nucleus only slightly stretched. The force is about 2 pN. At higher strains (regions IV and V), the IFs are stiffer (see force-strain curve in Figure 5.4), thus they resist more deformation and force the nucleus to be stretched. Therefore applied forces are much higher than at small strains (32 pN for region IV and 182 pN for region V).

Figure 5.8 shows snapshots of cell models with 40 IFs (region I), 20 Ifs (region II) and 10 IFs (region III) at an applied tensile strain of 48%. We observe that the nucleus is increasingly stretched as the IF density increases. At a given strain, the more IFs, the stiffer the cell is, then the higher is the stress and thus the more stretched is the nucleus.

5.3 Conclusion

The results reported in this Chapter illustrate the IF contribution to cell stiffness and deformation and in particular their role of maintaining cell integrity at large deformations. Note that on the lateral sides of the cell, IFs do not resist cell compression since their bending stiffness (or persistence length) is very low. This illustrates a limitation of our cell model that does not include microtubules (MTs). Indeed MTs have a much higher bending stiffness and

thus could make the cell resist compression more effectively. Moreover, the presence of actin microfilaments (MFs) would have made the cell stiffer for small strains.

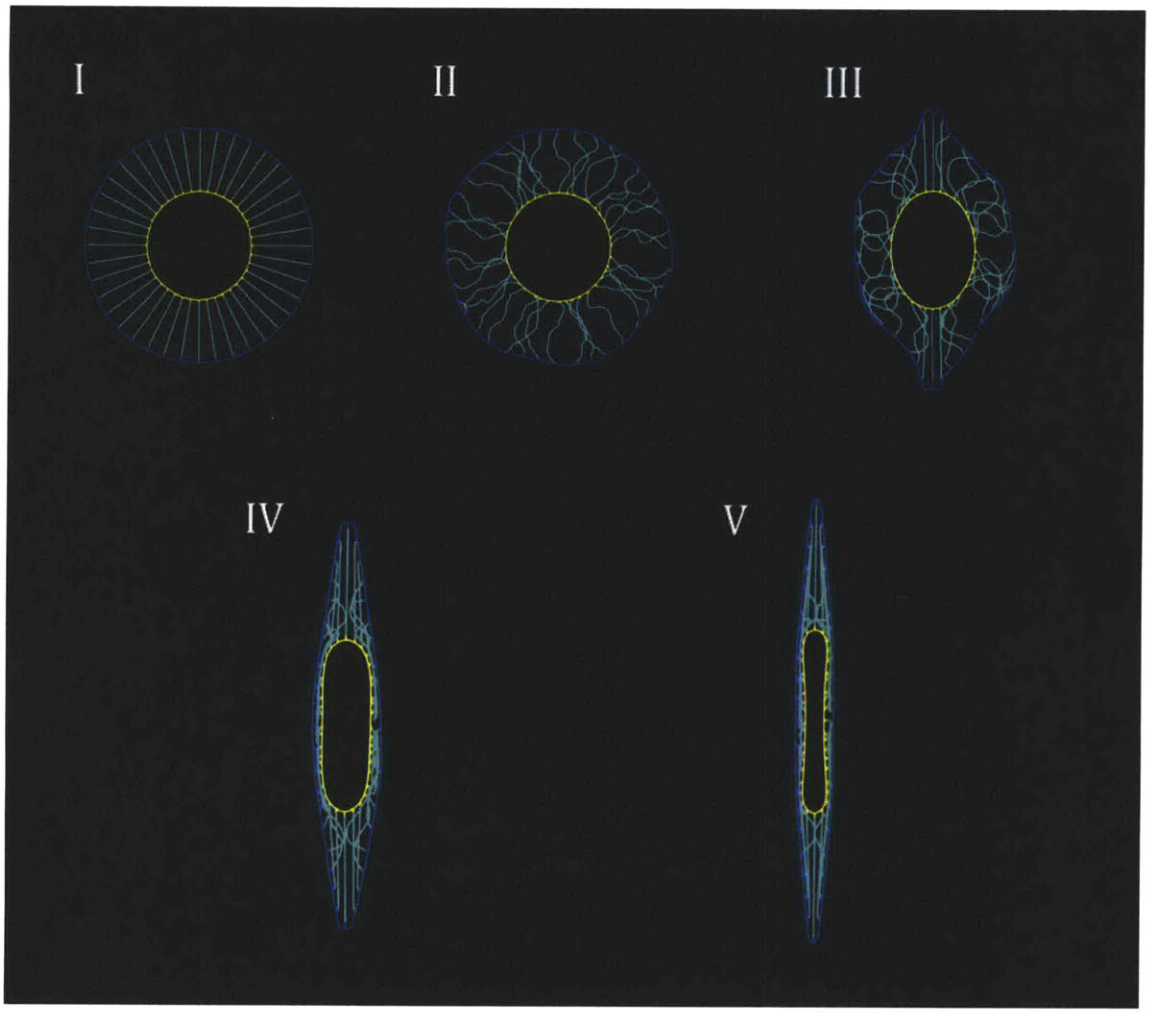

Figure 5.7. Snapshots of the cell model with 40 IFs at different applied tensile strains. Region I shows the initial model geometry. The system has 40 radial intermediate filaments uniformly peripherically distributed. Region II depicts the geometrical configuration of the model after equilibration (4000 ns). Note that IFs are slack and membranes have small fluctuations. Regions III, IV and V show snapshots of the cell model at various stretching strains (respectively 24%, 48% and 59%). As we pull the cell, the intermediate filament network is stretched in the areas between the pulled extremities and the nucleus, and compressed (bended) in the cell lateral areas. At small strains (region III), the IF network is deformed as described above and the nucleus only slightly streched. The force is about 2pN. At higher strains (regions IV and V), the IFs are stiffer (see force-strain curve in Figure 5.4), thus they resist more deformation and force the nucleus to be stretched. Therefore applied forces are much higher than at small strains (32pN for region IV and 182pN for region V). These results illustrate the IF contribution to cell stiffness and deformation and in particular their role of maintaining cell integrity at large deformations. Note that on the lateral sides of the cell, IFs do not resist cell compression since their bending stiffness (or persistence length) is very low. This illustrates a limitation of our cell model that does not include microtubules (MTs). Indeed MTs have a much higher bending stiffness and thus could make the cell resist compression more effectively. Moreover, the presence of actin microfilaments (MFs) would have made the cell stiffer for small strains.

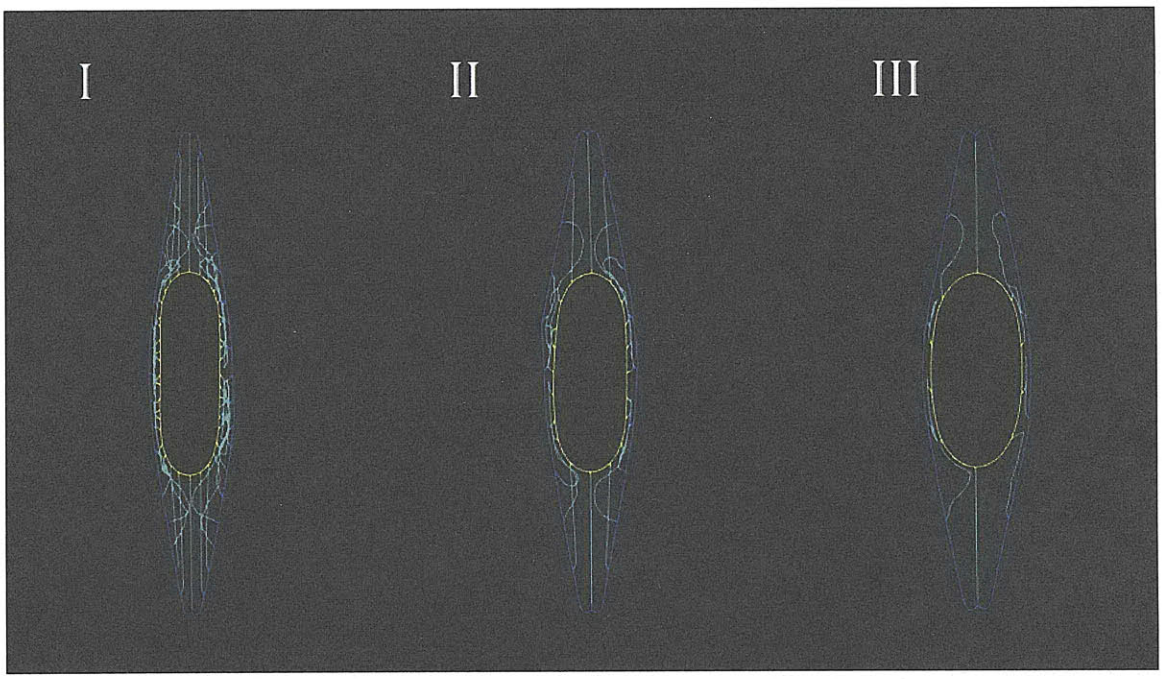

Figure 5.8. Snapshots of cell models with 40 IFs (region I), 20 Ifs (region II) and 10 IFs (region III) at an applied tensile strain of 48%. We observe that the nucleus is increasingly stretched as the IF density increases. At a given strain, the more IFs, the stiffer the cell is, then the higher is the stress and thus the more stretched is the nucleus.

6 Conclusion and future perspectives

6.1 Summary of main findings

In this Thesis, we focused on the deformation and failure mechanisms of hierarchical alpha-helical protein materials. We have developed, calibrated and applied two mesoscale computational models to describe respectively the deformation and failure mechanisms of AH protein arrangements and a eukaryotic cell. We summarize our main findings:

1. Discovered fundamental structure-process-property relationships in AH protein materials

   The computational model for AH protein arrangements is built in a way that one can control the structural parameters such as chemical bonding, adhesion energy, length and number of AH strands and thus measure their effects on properties such as strength. We have carried out a systematic variation of the parameters in the mesoscale model and studied their impacts on the computational results and compared with commonly used protein strength models. We summarize the main findings below:

   - The reference system provided us with the starting point for a systematic variation of all relevant parameters, to identify how strength properties depend on the details of the energy landscape and under which conditions the Bell model fails. By systematically varying the energy landscape parameters, we have shown that the Bell model is a reasonable approximation to predict the mechanical strength properties as long as the model parameters are in a certain range: E_b between 9-18 kcal/mol (weak bond energy barrier) and x_r value above x_b value. (see, e.g. Figure 4.10). Conversely, the Bell model breaks down when the model parameters are not in this range. We have also analyzed the dependence of the unfolding force on E_b and x_b (Figure 4.12) and confirmed the predictions from the Bell model.

   - We have shown that the properties of the second, unfolded state of a AH convolution do not strongly influence the strength properties, provided that x_r, the distance between the unfolded state and the transition state, is higher than x_b, the distance between the folded state and the transition state.(see, e.g. Figure 4.11).

We have also shown that the energetic properties of the second unfolded state do not influence the rupture strength.

- We have shown that under extremely small pulling velocities, the rupture strength of AH proteins approach an asymptotic value (Figure 4.15), where the strength does not depend on the pulling speed any more. For pulling speed above the critical rate, we have shown that the Bell model is a reasonable approximation.

- We have also reported an analysis of the effect of varying transducer stiffnesses on the strength properties, showing that stiffer transducers generally lead to increased strength properties (Figure 4.17).

- We have shown that under extremely small pulling speed, the model with change of bending stiffness is slightly stiffer than the model without this implementation (Figure 4.15).

- The results illustrate a very slight dependence of the rupture force on the bead mass (Figure 4.20).

- We have shown that interprotein sliding is a dominating mechanism that persists for a variety of geometries and realistic amino acid sequences (Figure 4.21). This suggests that alpha-helices remain largely intact during deformation of alpha-helix assemblies. Earlier experimental results on deformation of alpha-helical protein assemblies inside cells concur with this concept [94]. For very large adhesion values (in a range that does not correspond to adhesion properties associated with natural alpha-helices), our model predicts a change of deformation mechanisms where both sliding and unfolding occur concurrently. These strong adhesion values could be realized through the introduction of stronger interprotein bonding, e.g. via disulfide bonds that may provide a much greater shear resistance in excess of 1 nN. Future studies could explore this phenomenon.

- We have applied our mesoscale model to predict the strength of AH arrangements with different geometries and different lengths. For individual AH proteins, the study of length effect on strength shows two different regimes. For short AHs, we have shown a weakening effect as the molecule becomes longer. This could be explained by the larger number of serial coupled bonds, where failure of one bond

is sufficient to initiate failure of the entire system (Figure 4.23). For long AHs, we have shown the existence of a transition from the strength decrease regime to a strength asymptotic regime. An explanation could be that the deformation of longer systems is smoother than shorter systems because they have much more interactions and thus the peak of failure is averaged (Figure 4.24). This might be one of the reasons why short single alpha-helices are more prevalent in nature than long ones, as they provide increased strength against failure. For parallel AH proteins, we have shown that the length does not affect significantly the shear strength (Figure 4.27) and sliding is the prevalent deformation mechanism (versus unfolding) (Figures 4.21 and 4.26).

The applications of our coarse-grained model to structures of very different adhesion energies due to variations in amino acid sequences, lengths and arrangements illustrates the unique ability of our mesoscale model to describe how adhesion properties and chemical structure, geometrical structure (that is size, length, arrangement geometry) influence strength properties and mechanisms of deformation of protein arrangements. Earlier attempts of describing the deformation mechanisms of protein domains with "coarse" mesoscale models have not yet included such level of detail, and have thus not been capable of describing how adhesion, chemical structure, size, length, geometry and deformation mechanisms and strength are linked.

Individual, isolated AHs are rarely found in biology. However, the conclusions put forth here should be generically valid for a broader class of structural proteins. This is because the energy landscape structure (double-well potential) as presented in Figures 4.1-4.2 is not unique to alpha-helices. Rather, other protein filaments featuring serially arranged domains that unfold under strain can be described based on the modeling framework proposed here (with appropriate parameterization of the model parameters listed in Table 4.1). In light of this, the choice of AHs in the studies reported can be considered as a model system that helped us to elucidate the generic behavior of a general class of protein filaments. Specifically, the insight into the length-dependence of strength properties (Figures 4.23-4.24), the effects of pulling speed (Figure 4.15) and transducer stiffness (Figure 4.17) might be useful for the interpretation of experimental results and the comparison with theoretical and numerical models for many other protein filaments.

2. Developed multiscale models that predict elasticity, deformation and fracture of hierarchical assemblies of alpha-helical structures.

   We have developed, calibrated and validated two multiscale computational models that predict respectively the deformation and failure mechanisms of AH protein arrangements and a eukaryotic cell:

   - We have developed and calibrated (by fitting against full-atomistic MD results) a mesoscale model of AH protein arrangements, representing one convolution as a pair of mesoscale bead particles (Figure 4.2). This represents a reference system, which agrees well with the predictions from Bell's or Evans' models (Figures 4.4 and 4.5). The coarse-grained model enables us to simulate the dynamics of large systems over a large range of length- and time-scales and specific molecular properties of the constituting proteins. The model is capable of reaching time-scales of several microseconds and longer with a quantitative accuracy comparable with full atomistic MD simulations. Such relatively long simulations can be carried out within several days of computational time (on a single Intel Xeon CPU). In comparison, MD simulations of the dynamical behavior at fractions of microseconds can take weeks and months of computational time (even on a large parallelized simulation setup). Our coarse-grained model provides a considerable speedup while the model is still capable of describing the small- and large-deformation force-strain response characteristics as well as strength values quite accurately compared with atomistic simulations (see Figures 4.4-4.8). Earlier attempts of describing the deformation mechanisms of protein domains with "coarse" mesoscale models have not yet included such level of detail, and have thus not been capable of describing how adhesion, chemical structure, size, length, geometry and deformation mechanisms and strength are linked.

   - We developed a mesoscale description of a cell with nucleus (*i.e.* a eukaryotic cell). The cell model is set up not based on a rigorous hierarchical multiscale approach including atomistic features; rather, it is calibrated on a combination of experimental data and simulation data obtained on relatively big materials such as single intermediate filament or membranes that already are hierarchical structures. While this model is simplistic, it agrees well with experimental measurements (Figure 5.6) and enables us to gain qualitative insight into the deformation

behavior of cells (Figures 5.6-5.8) and illustrates how the multiscale approach developed in this Thesis can be used to describe more complex biological structures quite accurately compared with experiments.

3. Developed a model that could be used to improve the current understanding of pathological pathways linked to alpha-helical topologies such as muscular dystrophies.

   Through the simple mesoscale cell model developed here, we have shown that the cell model exhibits greater stiffening and greater stiffness as IF density increases and that these computational results are in good agreement with experimental data (Figure 5.6). Therefore we have shown that cytoskeletal intermediate filaments contribute to cell stiffness and deformation and thus they may play a significant role to maintain cell structural integrity in response to a given stress. As pointed out in [8], it is envisioned that the long-term potential impact of this work can be used to predict diseases in the context of diagnostic tools by measuring material properties rather than focusing on symptomatic chemical readings alone.

6.2 Current limitations and future perspectives

The mesoscale model for AH arrangements has some limitations that could be addressed in future work. Compared to full atomistic simulations, the coarse-grained model is much less detailed and may only give a rough description of deformation mechanisms. Figures 4.7 and 4.8 illustrate that, in contrast to atomistic simulations, the mesoscale model do not show initial unfolding before sliding. Further, although we have included the effects of variations of the amino acid sequence on the intermolecular adhesion, we have not yet explored the effect of sequence variations on their rupture properties. Based on the analysis of the range of rupture strengths of alpha-helical proteins (see error bar associated with the variation of H-bond strength in Figure 4.21), our results should hold for a variety of cases as the shearing strength is lower than the lower bound of the error bar. Choosing appropriate models for interatomic interactions provides a rather challenging and crucial step that remains subject of a very active discussion in the scientific community. A variety of different interatomic potentials are used in the studies of biological materials at different scales, and different types of protein structures may require the use of different atomistic models. A drawback of atomistic

simulations is the difficulty of analyzing results and the large computational resources necessary to perform the simulations. Due to computational limitations, MD simulations are restricted with respect to the timescales that can be reached, limiting overall time spans in such studies to tens of nanoseconds, or in very long simulation studies to fractions of microseconds. Therefore, many MD simulation results of dynamically stretching protein molecules, for instance, have been carried out at large deformation rates, exceeding several m/sec. For very large adhesion values (in a range that does not correspond to adhesion properties associated with natural alpha-helices), our model predicts a change of deformation mechanisms where both sliding and unfolding occur concurrently. These strong adhesion values could be realized through the introduction of stronger interprotein bonding, e.g. via disulfide bonds that may provide a much greater shear resistance in excess of 1 nN. Future studies could explore this phenomenon.

Moreover, individual, isolated AHs are rarely found in biology. Our mesoscale model for AH proteins cannot describe the formations of coiled coil or beta-sheet or tertiary structures as we observed it sometimes in full atomistic simulations and in particular for long alpha-helical strands. Further studies could focus on larger variations of timescales (e.g. to extend to even slower pulling speeds) and on applying the double-well potential to develop similar formulations for coiled-coil proteins, larger-level hierarchical protein structures or larger scale protein folds with tertiary structures. Models including the description of the transition between alpha-helix and beta-sheet structures could be also developed in the future.

The cell model have lots of limitations, in particular because it is simplistic and reductionist. First of all, when developing a 2D model to represent a 3D structure that does not present adequate symmetries or invariances, we may lose accuracy. Further, our model does not include the other two cytoskeletal elements which are the microtubules (MTs) and the microfilaments (MFs), which also play an important role in cell structural integrity. For instance, on the lateral sides of the stretched cell, IFs do not resist cell compression since their bending stiffness (or persistence length) is very low (Figure 5.6). In reality, MTs have a much higher bending stiffness and thus make the cell resist compression more effectively. Moreover, the presence of actin microfilaments (MFs) would have made the cell stiffer for small strains. Cell models including in MTs and MFs could be developed in the future. The ultimate goal would be to build a fully multiscale cell model by combining our two models (small and long scales) so that one can have a multiscale model that predicts cell deformation while taking into account the molecular structure as well as the hierarchical features, and thus

7 References

1. Astbury, W.T. and A. Street, *X-ray studies of the structures of hair, wool and related fibres. I. General.* Trans. R. Soc. Lond. A, 1931. **230**: p. 75-101.
2. Alberts, B., et al., *Molecular Biology of the Cell.* 2002, New York: Taylor & Francis.
3. Weiner, S. and H.D. Wagner, *The material bone: Structure mechanical function relations.* Annual Review Of Materials Science, 1998. **28**: p. 271-298.
4. Vincent, J.F.V., *Structural biomaterials*, ed. Anonymous. 1990, Princeton, N.J.: Princeton University Press. 244.
5. Fratzl, P. and R. Weinkamer, *Nature's hierarchical materials.* Progress in Materials Science, 2007. **52**(8): p. 1263-1334.
6. Buehler, M.J. and S. Keten, *Elasticity, strength and resilience: A comparative study on mechanical signatures of α-helix, β-sheet and tropocollagen domains.* Nano Research, 2008. **1**(1): p. 63-71.
7. Luo, L., et al., *Familial aggregation of age-related macular degeneration in the Utah population.* Vision Res, 2008. **48**(3): p. 494-500.
8. Buehler, M.J. and Y.C. Yung, *Deformation and failure of protein materials in physiologically extreme conditions and disease.* Nature Materials, 2009. **8**(3): p. 175-188.
9. Buehler, M.J. and T. Ackbarow, *Fracture mechanics of protein materials.* Materials Today, 2007. **10**(9): p. 46-58.
10. Engler, A.J., et al., *Matrix elasticity directs stem cell lineage specification.* Cell, 2006. **126**(4): p. 677-689.
11. Yung, Y.C., et al., *Cyclic tensile strain triggers a sequence of autocrine and paracrine signaling that regulate angiogenesis in human vascular cells.* P. Natl. Acad. Sci. USA. accepted for publication (in press).
12. Gautieri, A., et al., *Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils.* Biophysical Journal, 2009. **97**.
13. Suresh, S., et al., *Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria.* Acta Biomaterialia, 2005. **1**(1): p. 15-30.
14. Cross, S.E., et al., *Nanomechanical analysis of cells from cancer patients.* Nature Nanotechnology, 2007. **2**: p. 780–783.
15. Buehler, M.J., S. Keten, and T. Ackbarow, *Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture.* Progress in Materials Science, 2008 **53**: p. 1101-1241.
16. Courtney, T.H., *Mechanical behavior of materials.* 1990, New York, NY, USA: McGraw-Hill.
17. Broberg, K.B., *Cracks and Fracture.* 1990: Academic Press.
18. Hirth, J.P. and J. Lothe, *Theory of Dislocations.* 1982: Wiley-Interscience.
19. Buehler, M.J., *Atomistic modeling of materials failure.* 2008: Springer (New York).
20. Gruber, M. and A.N. Lupas, *Historical review: Another 50th anniversary - new periodicities in coiled coils.* Trends in Biochemical Sciences, 2003. **28**(12): p. 679-685.
21. Moir, R.D. and T.P. Spann, *The structure and function of nuclear lamins: implications for disease.* Cellular and Molecular Life Sciences, 2001. **58**(12-13): p. 1748-1757.
22. Wilson, K.L., M.S. Zastrow, and K.K. Lee, *Lamins and disease: Insights into nuclear infrastructure.* Cell, 2001. **104**(5): p. 647-650.

describes the deformation mechanisms and mechanical contributions of each hierarchical stage. Then the relationship between macroscopic properties and atomistic structures and hierarchical features of these biological protein materials would become fully understood.

Once such multiscale models would be developed, there is no doubt that they would open the door to applications with very significant payoff in many fields such as medicine, biological and material engineering.

23. Makarov, D.E., P.K. Hansma, and H. Metiu, *Kinetic Monte Carlo simulation of titin unfolding.* Journal Of Chemical Physics, 2001. **114**(21): p. 9663-9673.
24. Rathore, N., Q.L. Yan, and J.J. de Pablo, *Molecular simulation of the reversible mechanical unfolding of proteins.* Journal Of Chemical Physics, 2004. **120**(12): p. 5781-5788.
25. Bryson, J.W., et al., *Protein design - a hierarachical approach.* Science, 1995. **270**(5238): p. 935-941.
26. Kirshenbaum, K., R.N. Zuckermann, and K.A. Dill, *Designing polymers that mimic biomolecules.* Current Opinion in Structural Biology, 1999. **9**(4): p. 530-535.
27. Vashishta, P., R.K. Kalia, and A. Nakano, *Large-scale atomistic simulations of dynamic fracture.* Comp. in Science and Engrg., 1999: p. 56-65.
28. Rountree, C.L., et al., *Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular dynamics simulations.* Annual Rev. of Materials Research, 2002. **32**: p. 377-400.
29. Buehler, M.J. and H.J. Gao, *Dynamical fracture instabilities due to local hyperelasticity at crack tips.* Nature, 2006. **439**(7074): p. 307-310.
30. Buehler, M.J., F.F. Abraham, and H. Gao, *Hyperelasticity governs dynamic fracture at a critical length scale.* Nature, 2003. **426**: p. 141-146.
31. Buehler, M.J. and H. Gao, *Ultra large scale atomistic simulations of dynamic fracture* in *Handbook of Theoretical and Computational Nanotechnology, Edited by W. Schommers and A. Rieth.* 2006, American Scientific Publishers (ASP).
32. Buehler, M.J., A.C.T.v. Duin, and W.A. Goddard, *Multi-paradigm modeling of dynamical crack propagation in silicon using the ReaxFF reactive force field.* Phys. Rev. Lett. , 2006. **96**(9): p. 095505
33. Buehler, M.J., et al., *Threshold Crack Speed Controls Dynamical Fracture of Silicon Single Crystals.* Phys. Rev. Lett., 2007. **99**: p. 165502
34. Wang, W., et al., *Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions.* Annual Review Of Biophysics And Biomolecular Structure, 2001. **30**: p. 211-243.
35. Mackerell, A.D., *Empirical force fields for biological macromolecules: Overview and issues.* Journal of Computational Chemistry, 2004. **25**(13): p. 1584-1604.
36. Deniz, A.A., S. Mukhopadhyay, and E.A. Lemke, *Single-molecule biophysics: at the interface of biology, physics and chemistry.* Journal of the Royal Society Interface, 2008. **5**(18): p. 15-45.
37. Scheraga, H.A., M. Khalili, and A. Liwo, *Protein-folding dynamics: Overview of molecular simulation techniques.* Annual Review of Physical Chemistry, 2007. **58**: p. 57-83.
38. Seibert, M.M., et al., *Reproducible polypeptide folding and structure prediction using molecular dynamics simulations.* Journal Of Molecular Biology, 2005. **354**(1): p. 173-183.
39. Nelson, M.T., et al., *NAMD: A parallel, object oriented molecular dynamics program.* International Journal Of Supercomputer Applications And High Performance Computing, 1996. **10**(4): p. 251-268.
40. Ponder, J. and D. Case, *Force fields for protein simulations.* PROTEIN SIMULATIONS, 2003. **66**: p. 27-+.
41. Goddard, W.A., *A Perspective of Materials Modeling* in *Handbook of Materials Modeling*, S. Yip, Editor. 2006, Springer.

42. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
43. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - A Generic Force-Field For Molecular Simulations.* Journal Of Physical Chemistry, 1990. **94**(26): p. 8897-8909.
44. Rappe, A.K., et al., *Uff, A Full Periodic-Table Force-Field For Molecular Mechanics And Molecular-Dynamics Simulations.* Journal Of The American Chemical Society, 1992. **114**(25): p. 10024-10035.
45. Pearlman, D.A., et al., *Amber, A Package Of Computer-Programs For Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics And Free-Energy Calculations To Simulate The Structural And Energetic Properties Of Molecules.* Computer Physics Communications, 1995. **91**(1-3): p. 1-41.
46. Tirion, M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* PHYSICAL REVIEW LETTERS, 1996. **77**(9): p. 1905-1908.
47. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* PHYSICAL REVIEW LETTERS, 1997. **79**(16): p. 3090-3093.
48. HAYWARD, S. and N. GO, *COLLECTIVE VARIABLE DESCRIPTION OF NATIVE PROTEIN DYNAMICS.* ANNUAL REVIEW OF PHYSICAL CHEMISTRY, 1995. **46**: p. 223-250.
49. Tozzini, V., *Coarse-grained models for proteins.* CURRENT OPINION IN STRUCTURAL BIOLOGY, 2005. **15**(2): p. 144-150.
50. West, D.K., et al., *Mechanical resistance of proteins explained using simple molecular models.* Biophysical journal, 2006. **90**(1): p. 287-297.
51. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics.* PHYSICAL REVIEW LETTERS, 2008. **1**(9): p. -.
52. Sulkowska, J.I. and M. Cieplak, *Mechanical stretching of proteins - a theoretical survey of the Protein Data Bank.* Journal of Physics-Condensed Matter, 2007. **19**(28): p. -.
53. Bathe, M., *A finite element framework for computation of protein normal modes and mechanical response.* PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS, 2008. **70**(4): p. 1595-1609.
54. Bahar, I., M. Kaplan, and R.L. Jernigan, *Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches.* Proteins-Structure Function And Genetics, 1997. **29**(3): p. 292-308.
55. Nguyen, H. and C. Hall, *Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides.* PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 2004. **101**(46): p. 16180-16185.
56. Nguyen, H. and C. Hall, *Spontaneous fibril formation by polyalanines; Discontinuous molecular dynamics simulations.* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, 2006. **128**(6): p. 1890-1901.
57. Freddolino, P., A. Arkhipov, and K. Schulten, *Coarse-grained molecular dynamics simulations of rotation-induced structural transitions in the bacterial flagellum.* BIOPHYSICAL JOURNAL, 2007: p. 142A-142A.
58. Buehler, M.J., *Nature designs tough collagen: Explaining the nanostructure of collagen fibrils.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(33): p. 12285-12290.
59. Buehler, M., *Molecular nanomechanics of nascent bone: fibrillar toughening by mineralization.* NANOTECHNOLOGY, 2007. **18**(29): p. -.

60. Bell, G.I., *Models for the specific adhesion of cells to cells.* Science, 1978. **200**(4342): p. 618-627.
61. Hanggi, P., P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after Kramers.* Rev. Mod. Phys, 1990. **62**(2): p. 251–341.
62. Zhurkov, S.N., *Kinetic concept of the strength of solids. Int.* Journal of Fracture Mechanics, 1965. **1**: p. 311-323.
63. Evans, E. and K. Ritchie, *Dynamic strength of molecular adhesion bonds.* Biophysical Journal, 1997. **72**(4): p. 1541-1555.
64. Hyeon, C. and D. Thirumalai, *Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments.* Journal of Physics, Condensed Matter, 2007. **19**(11): p. 113101.
65. Seifert, U., *Rupture of multiple parallel molecular bonds under dynamic loading.* Physical Review Letters, 2000. **84**(12): p. 2750-2753.
66. Seifert, U., *Dynamic strength of adhesion molecules: Role of rebinding and self-consistent rates.* Europhysics Letters, 2002. **58**(5): p. 792-798.
67. Evans, E., *P ROBING THE R ELATION B ETWEEN F ORCE-L IFETIME-AND C HEMISTRY IN S INGLE M OLECULAR B ONDS.* Annual Reviews in Biophysics and Biomolecular Structure, 2001. **30**(1): p. 105-128.
68. Hummer, G. and A. Szabo, *Kinetics from Nonequilibrium Single-Molecule Pulling Experiments.* Biophysical Journal, 2003. **85**(1): p. 5-15.
69. Walton, E.B., S. Lee, and K.J. Van Vliet, *Extending Bell's Model: How Force Transducer Stiffness Alters Measured Unbinding Forces and Kinetics of Molecular Complexes.* Biophysical Journal, 2008. **94**(7): p. 2621.
70. Zwanzig, R., *Diffusion in a Rough Potential.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(7): p. 2029-2030.
71. Erdmann, T. and U.S. Schwarz, *Stability of Adhesion Clusters under Constant Force.* Physical Review Letters, 2004. **92**(10): p. 108102.
72. Erdmann, T. and U.S. Schwarz, *Bistability of Cell-Matrix Adhesions Resulting from Nonlinear Receptor-Ligand Dynamics.* Biophysical Journal, 2006. **91**(6): p. L60.
73. Erdmann, T. and U.S. Schwarz, *Stability of adhesion clusters under constant force.* Physical Review Letters, 2004. **92**(10): p. 4.
74. Rief, M., J.M. Fernandez, and H.E. Gaub, *Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility.* Physical Review Letters, 1998. **81**(21): p. 4764-4767.
75. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics.* Physical Review Letters, 2008. **100**(9): p. 4.
76. Dietz, H. and M. Rief, *Elastic Bond Network Model for Protein Unfolding Mechanics.* Physical Review Letters, 2008. **100**(9): p. 98101.
77. Ackbarow, T., et al., *Hierarchies, multiple energy barriers and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains.* P. Natl. Acad. Sci. USA, 2007. **104**: p. 16410-16415
78. Bell, G.I., *Models for Specific Adhesion of Cells to Cells.* Science, 1978. **200**(4342): p. 618-627.
79. Buehler, M.J., *Hierarchical chemo-nanomechanics of stretching protein molecules: Entropic elasticity, protein unfolding and molecular fracture.* Journal of Mechanics of Materials and Structures, 2007. **2**(6): p. 1019-1057.
80. Karcher, H., et al., *A coarse-grained model for force-induced protein deformation and kinetics.* Biophysical Journal, 2006. **90**(8): p. 2686-2697.

81. Bertaud, J., Z. Qin, and M.J. Buehler, *Atomistically informed mesoscale model of alpha-helical protein domains.* International Journal for Multiscale Computational Engineering, 2009. 7(3): p. 237-250.
82. Ackbarow, T. and M.J. Buehler, *Superelasticity, energy dissipation and strain hardening of vimentin coiled-coil intermediate filaments: Atomistic and continuum studies.* Journal of Materials Science, 2007 **42**(21): p. 8771-8787.
83. Keten, S. and M.J. Buehler, *Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates.* Phys. Rev. Lett., 2008. **100**: p. 198301
84. Lazaridis, T. and M. Karplus, *Effective energy function for proteins in solution.* Proteins-Structure Function And Genetics, 1999. **35**(2): p. 133-152.
85. Lazaridis, T. and M. Karplus, *"New view" of protein folding reconciled with the old through multiple unfolding simulations.* Science, 1997. **278**(5345): p. 1928-1931.
86. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal Of Molecular Graphics, 1996. **14**(1): p. 33.
87. Husson, J. and F. Pincet, *Analyzing single-bond experiments: influence of the shape of the energy landscape and universal law between the width, depth, and force spectrum of the bond.* Phys Rev E Stat Nonlin Soft Matter Phys, 2008. 77(2 Pt 2): p. 026108.
88. Tshiprut, Z., Y. Klafter, and M. Urbakh, *Single-Molecule Pulling Experiments: When the Stiffness of the Pulling Device Matters.* Biophys J, 2008.
89. Ackbarow, T., S. Keten, and M.J. Buehler, *Multi-time scale strength model of alpha-helical protein domains.* Journal of Physics: Condensed Matter, 2009. **21**: p. 035111.
90. Dudko, O.K., et al., *Beyond the conventional description of dynamic force spectroscopy of adhesion bonds.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(20): p. 11378-11381.
91. Friddle, R.W., *Unified model of dynamic forced barrier crossing in single molecules.* Physical Review Letters, 2008. **100**(13): p. -.
92. Tshiprut, Z., J. Klafter, and M. Urbakh, *Single-molecule pulling experiments: When the stiffness of the pulling device matters.* Biophysical Journal, 2008. **95**(6): p. L42-L44.
93. Qin, Z., J. Zou, and X.Q. Feng, *Influence of Water on the Frequency of Carbon Nanotube Oscillators.* Journal Of Computational And Theoretical Nanoscience, 2008. **5**(7): p. 1403-1407.
94. Simon Penel, R.G.M., Russell J. Mortishire-Smith and Andrew J. Doig, *Periodicity in alpha-helix lengths and C-capping preferences.* Journal of Molecular Biology, 1999. **293**(5): p. 1211-1219.
95. Qi, H.J., C. Ortiz, and M.C. Boyce, *Mechanics of biomacromolecular networks containing folded domains.* Journal of Engineering Materials and Technology-Transactions of the Asme, 2006. **128**(4): p. 509-518.
96. Cheng, Q.H., et al., *A computational modeling for micropipette-manipulated cell detachment from a substrate mediated by receptor-ligand binding.* Journal Of The Mechanics And Physics Of Solids, 2009. **57**(2): p. 205-220.
97. Tang, Y.Y., et al., *Mechanosensitive channels: Insights from continuum-based simulations.* Cell Biochemistry and Biophysics, 2008. **52**(1): p. 1-18.
98. Ingber, D.E., *Tensegrity I. Cell structure and hierarchical systems biology.* Journal of Cell Science, 2003. **116**(7): p. 1157-1173.
99. Ackbarow, T., et al., *Alpha-Helical Protein Networks are Self Protective and Flaw Tolerant.* PLoS ONE, 2009. **4**(6): p. e6015.
100. Mucke, N., et al., *Assessing the flexibility of intermediate filaments by atomic force microscopy.* Journal of Molecular Biology, 2004. **335**(5): p. 1241-1250.

101. Fudge, D.S., et al., *The mechanical properties of hydrated intermediate filaments: Insights from hagfish slime threads.* Biophysical Journal, 2003. **85**(3): p. 2015-2027.
102. Ackbarow, T., et al., *Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**: p. 16410-16415.
103. Qin, Z., L. Kreplak, and M.J. Buehler, *Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments.* PLoS ONE, 2009. in submission.
104. Fudge, D.S., et al., *The mechanical properties of hydrated intermediate filaments: insights from hagfish slime threads.* Biophys J, 2003. **85**(3): p. 2015-27.
105. Waugh, R. and E.A. Evans, *Thermoelasticity of Red Blood-Cell Membrane.* Biophysical Journal, 1979. **26**(1): p. 115-131.
106. Mohandas, N. and E. Evans, *Rheological and Adherence Properties of Sickle Cells - Potential Contribution to Hematologic Manifestations of the Disease.* Sickle Cell Disease /, 1989. **565**: p. 327-337
107. Zhelev, D.V., D. Needham, and R.M. Hochmuth, *Role of the Membrane Cortex in Neutrophil Deformation in Small Pipettes.* Biophysical Journal, 1994. **67**(2): p. 696-705.
108. Wang, N. and D. Stamenovic, *Contribution of intermediate filaments to cell stiffness, stiffening, and growth.* American Journal of Physiology-Cell Physiology, 2000. **279**(1): p. C188-C194.